

This Page Is Inserted by IFW Operations
and is not a part of the Official Record

BEST AVAILABLE IMAGES

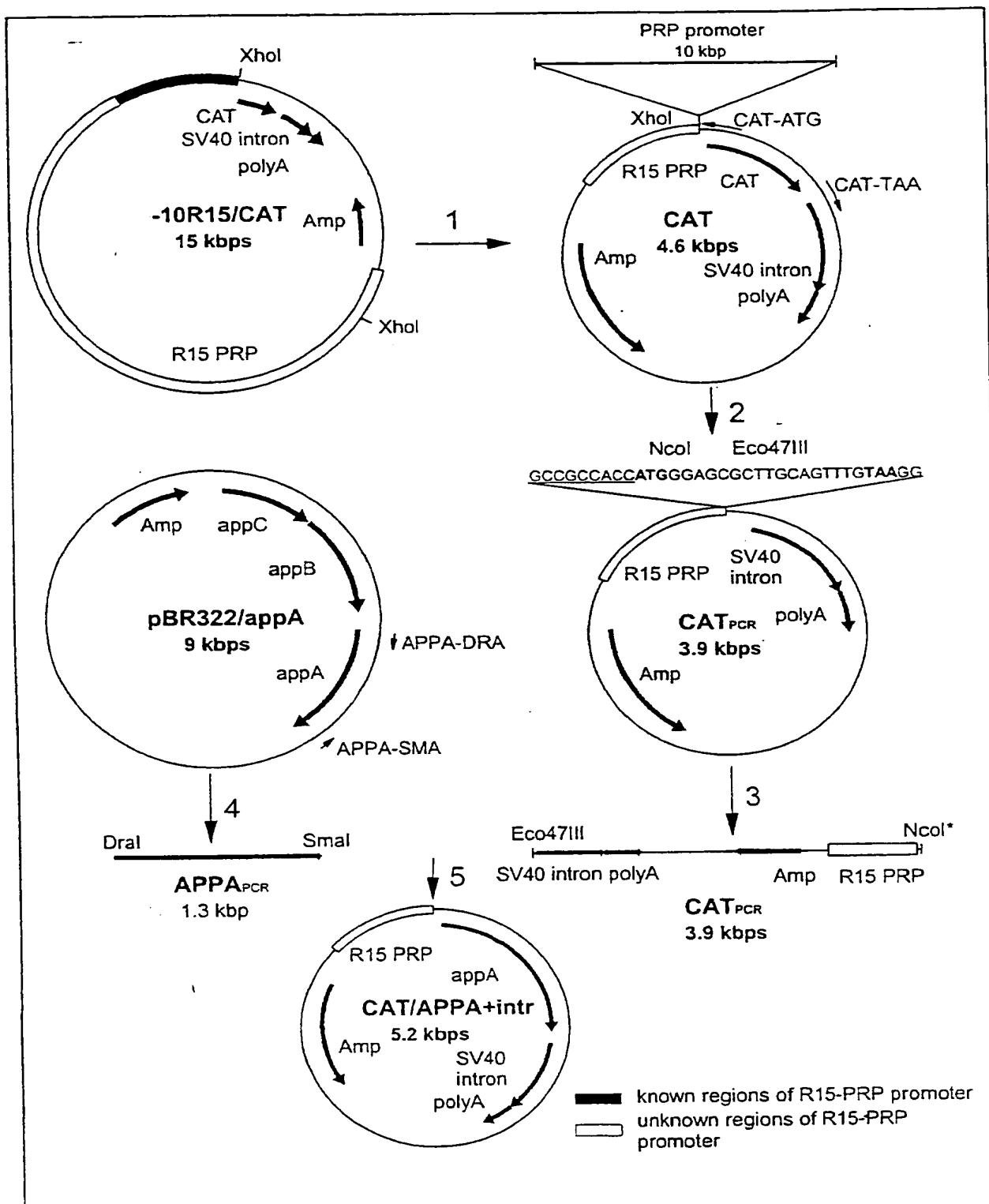
Defective images within this document are accurate representations of the original documents submitted by the applicant.

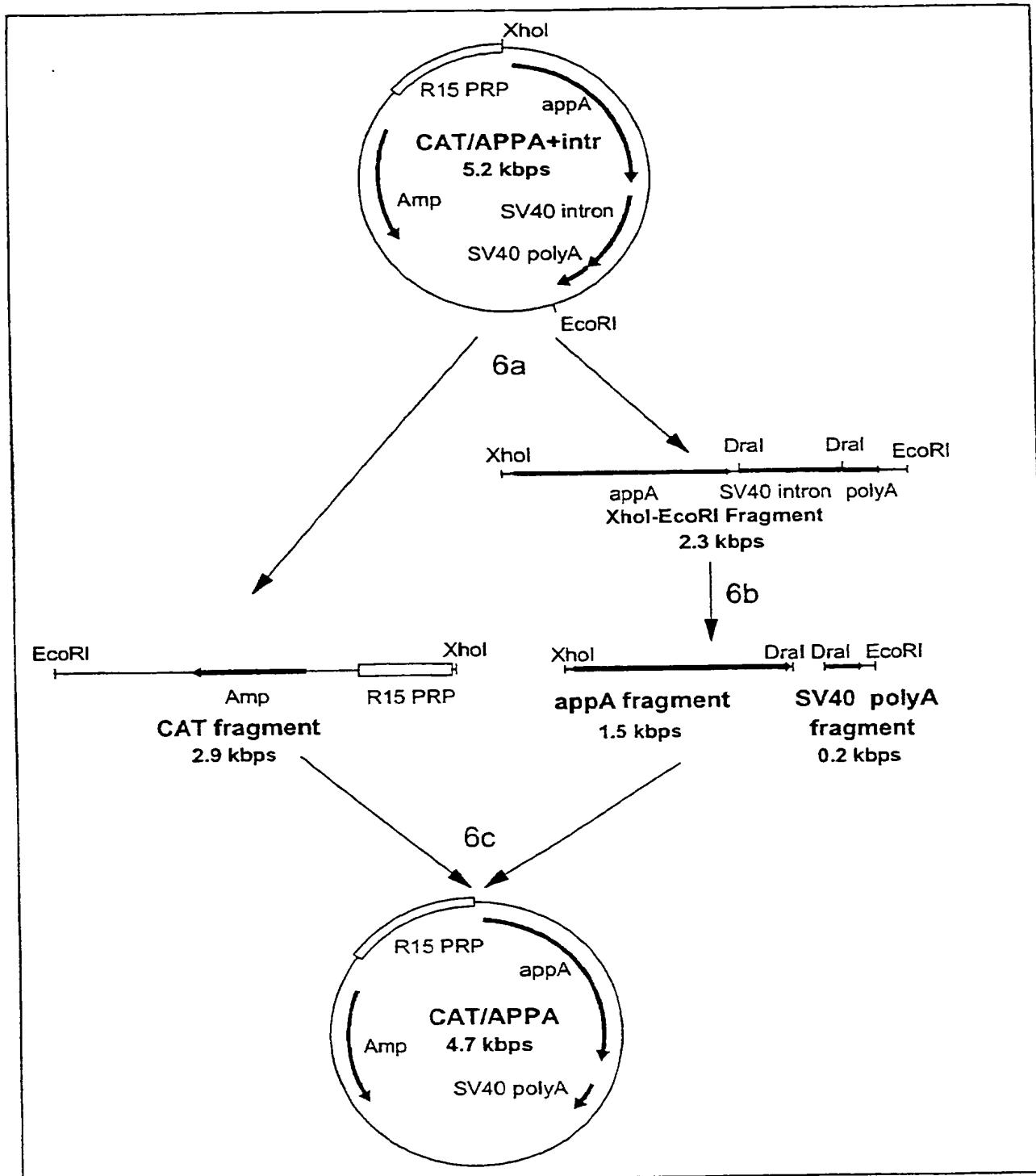
Defects in the images may include (but are not limited to):

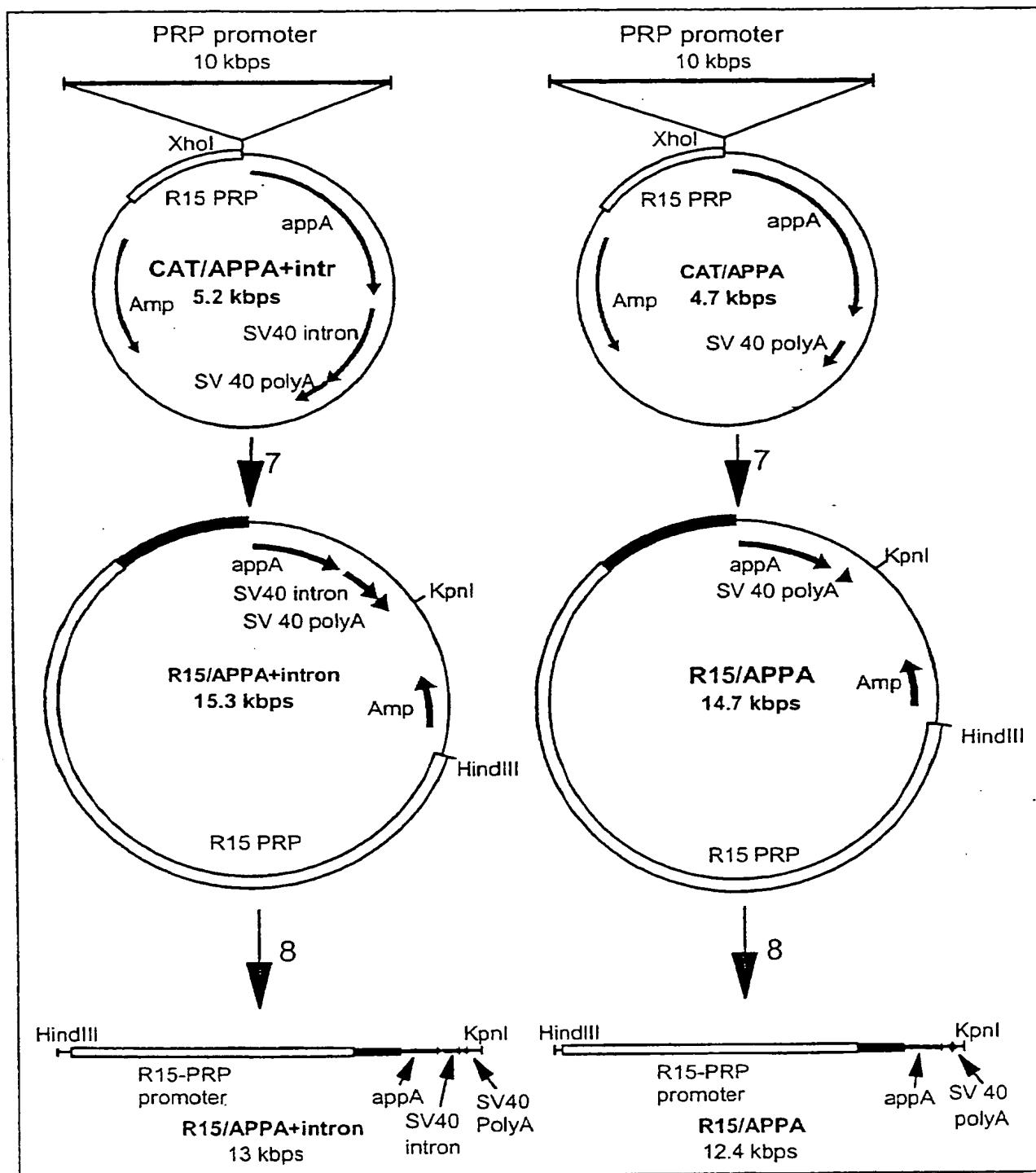
- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

**As rescanning documents *will not* correct images,
please do not report the images to the
Image Problem Mailbox.**

**Figure 1**

Figure 1 (continued)

**Figure 1 (continued)**

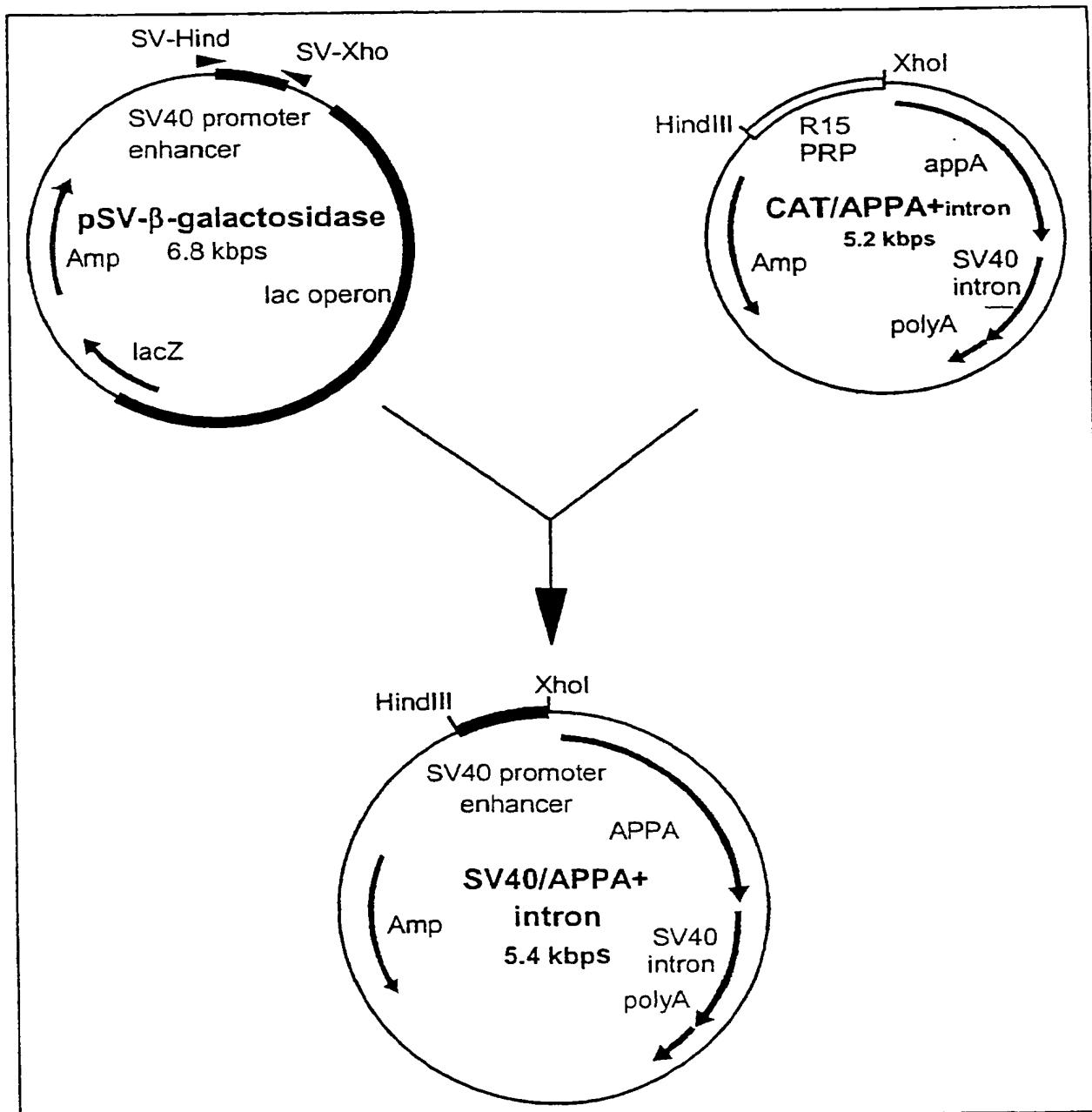
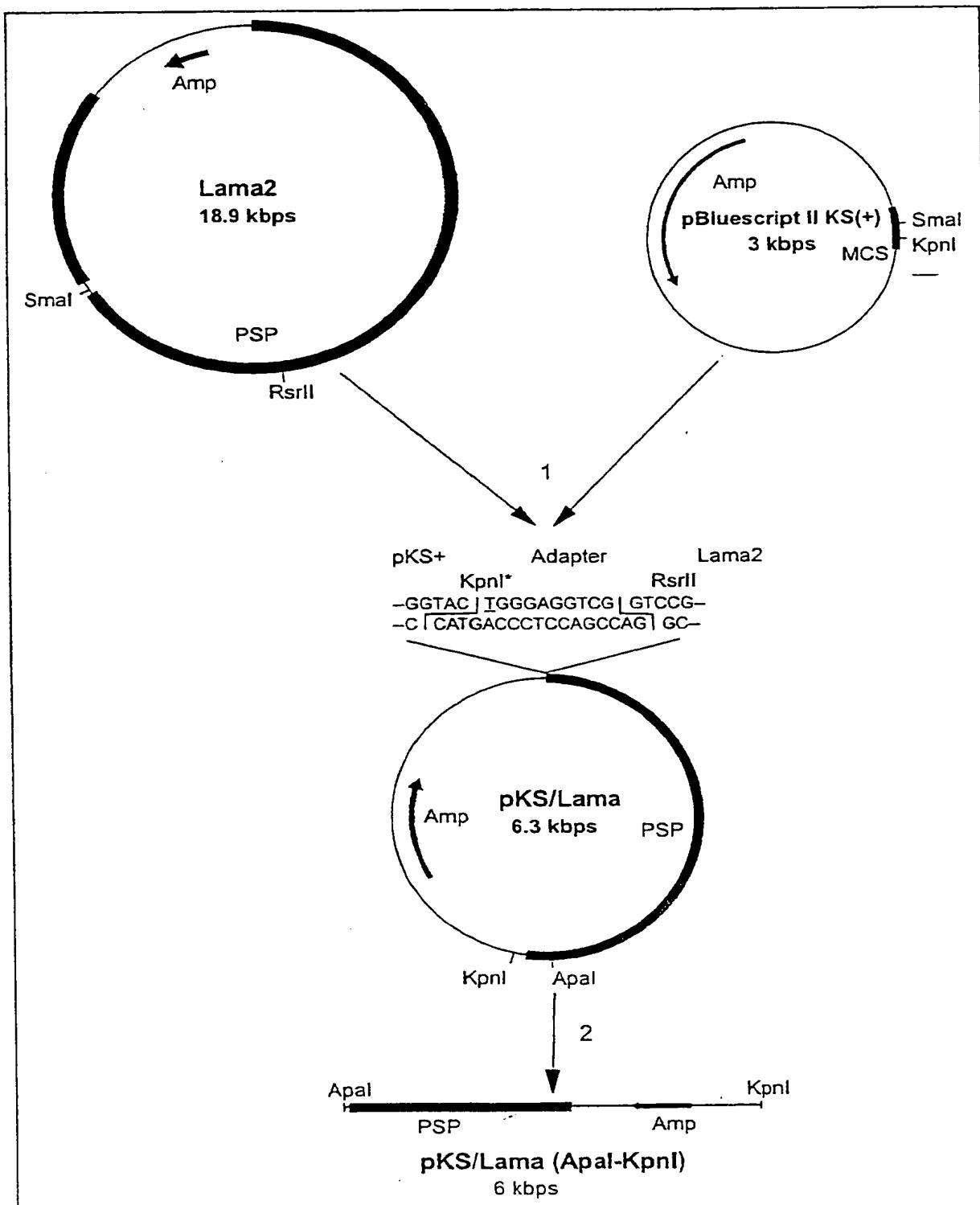
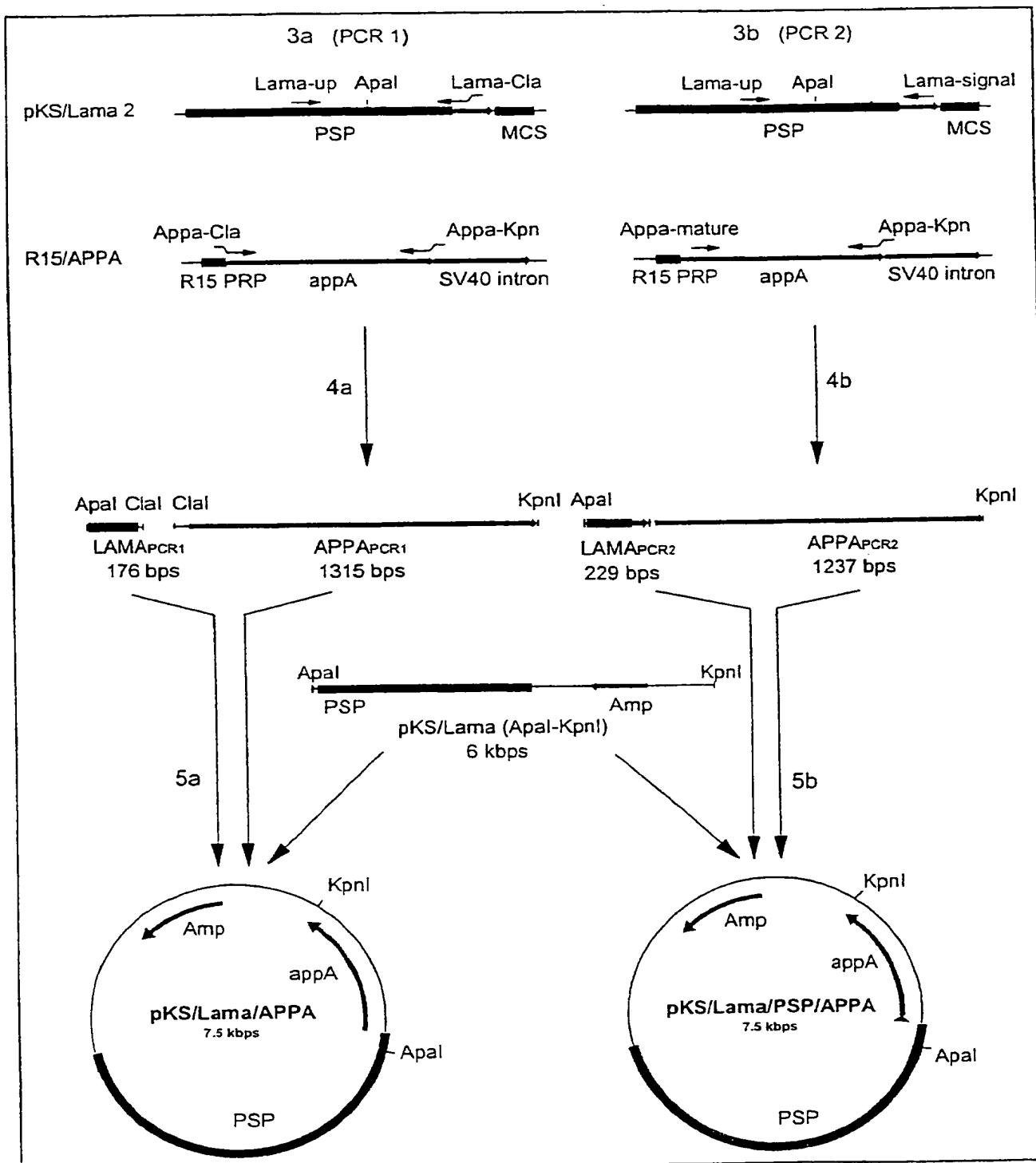


Figure 2

**Figure 3**

**Figure 3 (continued)**

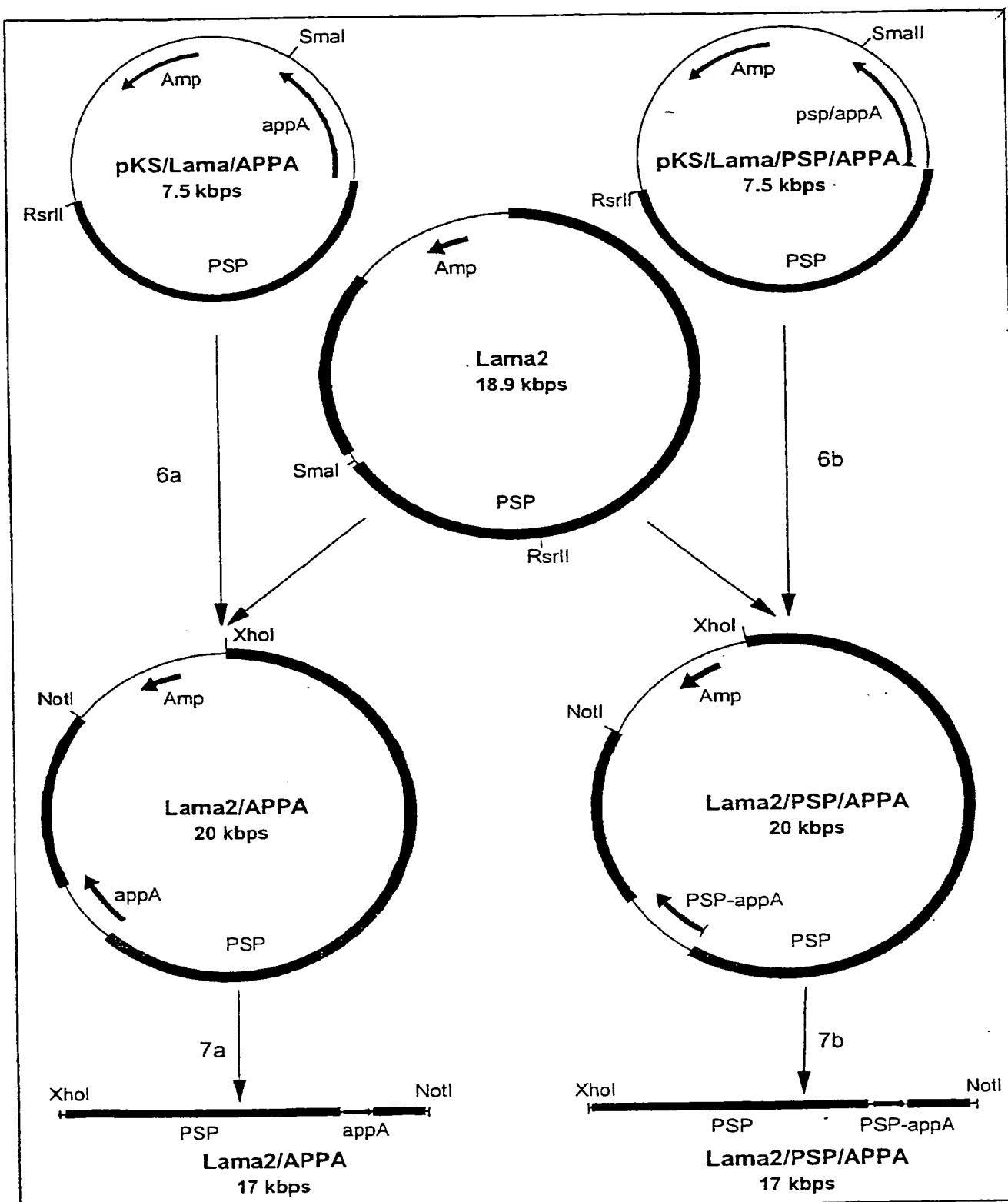


Figure 3 (continued)

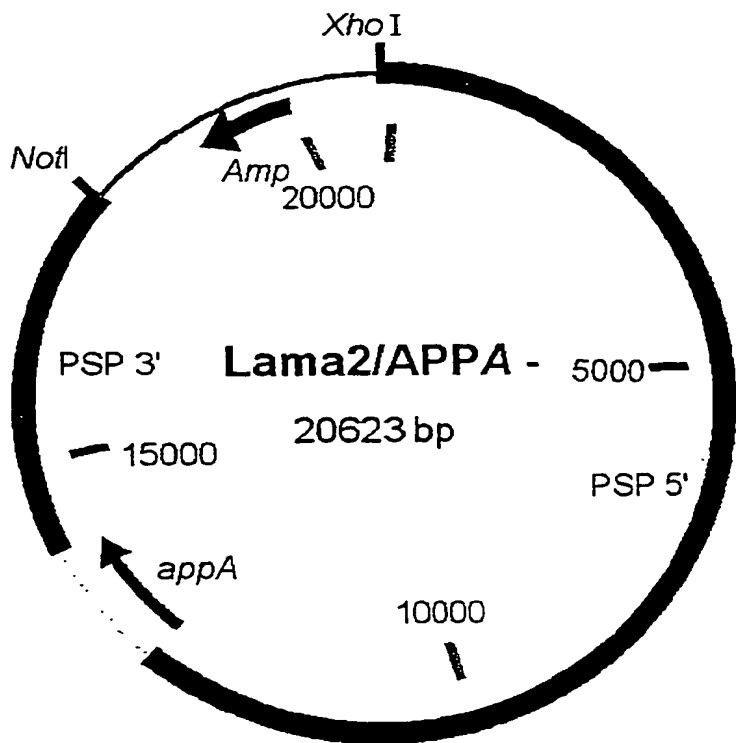


Figure 4. Schematic diagram of the Lama2/APPA construct.

Figure 5. The nucleic acid sequence of the Lama2/APPA plasmid (SEQ ID NO: 1)

LOCUS Lama-appA 20623 bp DNA CIRCULAR SYN 17-JAN-2000
 DEFINITION Lama 2/APPA transgenic construct
 ACCESSION Lama 2-appA,
 KEYWORDS parotid secretory protein; acid glucose-1-phosphatase; appA
 gene;
 periplasmic phosphoanhydride phosphohydrolase; artificial
 sequence;
 cloning vector
 REFERENCE 1 (bases 1 to 20623)
 AUTHORS Golovan, S., Forsberg, C.W., Phillips, J.
 JOURNAL Unpublished.
 FEATURES
 DEFINITION M. musculus Psp gene for parotid secretory protein.
 ACCESSION X68699
 VERSION X68699.1 GI:53809
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 3777 to 5332)
 AUTHORS Svendsen, P., Laursen, J., Krogh-Pedersen, H. and Hjorth, J.P.
 TITLE Novel salivary gland specific binding elements located in the PSP
 proximal enhancer core
 JOURNAL Nucleic Acids Res. 26 (11), 2761-2770 (1998)
 MEDLINE 98256451
 REFERENCE 2 (bases 7147 to 12653; 13952 to 17731)
 AUTHORS Mikkelsen, T.R.
 TITLE Direct Submission
 JOURNAL Submitted (07-OCT-1992) T.R. Mikkelsen, Department of Molecular
 Biology, University of Aarhus, CF Mollers Alle 130, 8000
 Aarhus, DENMARK
 REFERENCE 3 (bases 7147 to 12653; 13952 to 17731)
 AUTHORS Laursen, J., Hjorth, J.P.
 TITLE A cassette for high-level expression in the mouse salivary glands.
 JOURNAL Gene 1997 Oct 1;198(1-2):367-72
 MEDLINE 9370303

FEATURES Location/Qualifiers
 source 1..to 12653; 13952 to 17731
 /organism="Mus musculus"
 /strain="C3H/As"
 /db_xref="taxon:10090"
 /chromosome="2"
 /map="Estimate: 69 cM from centromere"
 /clone="Lambda YP1, Lambda YP3, Lambda YP7"
 /clone_lib="Lambda-PHAGE (Lambda L47.1)"
 /germline
 /note="Allele: b"
 misc_feature 3777-5332
 /gene="PSP"
 /function="salivary gland specific positive acting
 regulatory region"
 enhancer 7147..8724
 /evidence=experimental
 exon 11778..11824
 /gene="Psp"
 /note="exon a"
 /number=1
 /evidence=experimental
 exon 12626.. 14190
 /gene="Psp"
 /note="exon b fused with exons h and i"
 misc_feature 12644-12652

Figure 5 (continued):

```

        /function=" consensus sequence for initiation in higher
        eukaryotes "
misc_feature 13952..13965
        /function=" M13mp18 polylinker"

DEFINITION  E. coli periplasmic phosphoanhydride phosphohydrolase (appA) gene.

ACCESSION      M58708 L03370 L03371 L03372 L03373 L03374 L03375
VERSION        M58708.1 GI:145283
SOURCE          Escherichia coli DNA.
ORGANISM        Escherichia coli
                Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
                Escherichia.

REFERENCE      1 (bases 12653..13951)
AUTHORS        Dassa, J., Marck, C. and Boquet, P.L.
TITLE          The complete nucleotide sequence of the Escherichia coli gene appA
                reveals significant homology between pH 2.5 acid phosphatase
                and glucose-1-phosphatase
JOURNAL        J. Bacteriol. 172 (9), 5497-5500 (1990)
MEDLINE        90368616

FEATURES
Source          Location/Qualifiers
                12653..13951
                /organism="Escherichia coli"
                /db_xref="taxon:562"
sig_peptide     12653..12718
/gene="appA"
CDS12653       13951
                /gene="appA"
                /standard_name="acid phosphatase/phytase"
                /transl_table=11
                /product="periplasmic phosphoanhydride phosphohydrolase"
                /protein_id="AAA72086.1"
                /db_xref="GI:145285"

/translation="MKAILIPFLSLLIPLTPQSAFAQSEPELKLESVVIVSRHGVRAP
TKATQLMQDVTPDAWPTWPVKGWLTPRGGELIAYLGHYQRQRLVADGLLAKKGCPQS
GQVAAIADVDERTRKTGEAFAAGLAPDCAITVHTQADTSSPDPLFNPLKTGVCOLDNA
NVTDAILSRAGGSIAFDGHRQTAFRELERVLFNPQSNLCLKREKQDESCSLTQALPS
ELKVSADNVSLTGAVSLSMLTEIFLLQQAQGMPEPGWGRITDSHQWNTLLSLHNAQF
YLLQRTPEVARSRATPLLDLTIKTAITPHPPQKQAYGVTLPTSVLFIAQHDTNLANLGG
ALELNWTLPGQPDNTPPGGELEVFERWRRLSDNSQWIQVSLVFQTLQQMRDKTPLSLNT
PPGEVKLTLAGCEERNAQGMCSLAGFTQIVNEARI PACSL"
mat_peptide     12719 13948
                /gene="appA"
                /product="periplasmic phosphoanhydride phosphohydrolase"

mutation        replace(12659.. 12661,"gcg changed to gcc")
                /gene="appA"
                /standard_name="A3 mutant"
                /note="created by site directed mutagenesis"
                /citation=[3]
                /phenotype="silent mutation"
mutation        replace(13934..13936, " ccg changed to ccc")
                /gene="appA"
                /standard_name=" P428 mutant"
                /note="created by site directed mutagenesis"
                /citation=[3]
                /phenotype=" silent mutation "
mutation        replace(13937..13939, " gcg changed to gct")
                /gene="appA"
                /standard_name=" A429 mutant"
                /note="created by site directed mutagenesis"

```

Figure 5 (continued):

```

/citation=[3]
/phenotype=" silent mutation "

DEFINITION pBluescript II KS(+) vector DNA,
ACCESSION X52327
VERSION X52327.1 GI:58061
KEYWORDS artificial sequence; cloning vector; expression vector; vector.
SOURCE synthetic construct.
ORGANISM synthetic construct
REFERENCE 1 (bases 17732 to 20623)
AUTHORS Thomas, E.A.
TITLE Direct Submission
JOURNAL Submitted (20-FEB-1990) Thomas E.A., Stratagene Cloning
Systems, 11099 North Torney Pines Rd., La Jolla, CA 92037, USA
REFERENCE 2 (bases 17732 to 20623)
AUTHORS Short, J.M., Fernandez, J.M., Sorge, J.A. and Huse, W.D.
TITLE Lambda ZAP: a bacteriophage lambda expression vector with in
vivo excision properties
JOURNAL Nucleic Acids Res. 16 (15), 7583-7600 (1988)
MEDLINE 88319944
REFERENCE 3 (bases 17732 to 20623)
AUTHORS Alting-Mees, M.A. and Short, J.M.
TITLE pBluescript II: gene mapping vectors
JOURNAL Nucleic Acids Res. 17 (22), 9494 (1989)
MEDLINE 90067967
FEATURES Location/Qualifiers
Source 17732 to 20623
/organism="synthetic construct"
/db_xref="taxon:32630"
CDS complement (18967..19827)
/gene="Amp"
/product="b-lactamase"

```

BASE COUNT 5449 a 4847 c 4902 g 5424 t
 ORIGIN

```

1 TCGAGAGTAT CTTTGTCAAGC TGTGCCTCCA ACAAAAGGGT ACTGTTGCC ACATAGAAAG
 61 ATCTAAACTA ATTAATTAAAT CCCTCACCCG CAAATCTTC AGTCACTAAG TTAGCACGAT
121 TGTTGAACAA GTTCTCCAAA GGAGAGATAC AGATGAGTGC GTATAGGGTG GACCTGGCTG
181 CTGAGGAGAC ACCTGCATCT GACTAAGAAG AGCCACGGTG TTAGTTGAAT GGTGTGGAGT
241 AGGGTGGTTC TGTGGGACAG TAGAAAATCG AGAGGCATGT GCGTTTAGT GAACTGATGG
301 AAGCTACCCC AAACGACAGA GATTGTCAGT CAGGCCAATC CGTTTCGAGT TTGATGGCA
361 GCCGGACAGT GAGACAGACA CACCTACTCA GTTGGAGGAA GGATGAGAAC AATGCCAGC
421 AGGGATTGAG AGACCCTGAC AGGCGCAAGG CCCTAACACA CACACCTACC ACCTCACTTG
481 ACAAAAGCTGC CAAAGACCAA AGACTTGTTC TCCATTAGAA ATGACAGCTG GCTTGACCCG
541 ACAGCATAAT AAGCAGAGTG TACTCTGATT GGAGAACTTT AATGTGTTTC ATTCACTATT
601 ATAAAAGGAC AGTATTACAG ATTTTGTGT ACACTGCTGT TACATGTGGG GCAGTGTGTC
661 TTTAAGTAGG GTAAAGTACT CTTTAAAAAT GGGTCTAGA TATTTTTTCC TTTAACTCAA
721 GTCTCTTACT GTTTAAATGA TTTTTTATTT GTTTAATATG GAGGAAAAG AAGCTAAAT
781 GGACAATATA TATTTAGAGA AAGATGGTTA GCTGTCAGAA AAATATGCAA ATCAAAATCA
841 CACCAAGACT GCAGCACACC CCTGTCAGAT GGCTGTGATC AAGAAAATAA ATGACAATGA
901 GTGGTGGTGA AGATGTAATCA AAGGGAAACA CACACACACA CACACACACA
961 CACACTGGAG CAACCACGT GGAAATCAGT ATGAATGGTC CTCAAAACCC TGAAGATAGA
1021 GCGGGGCGTG GTGGCATACA CTTTTATTC CAGCACTGGG GAGGCAGAGG CAGGTGGATC
1081 TCTGAGTTC AGGGCAGCT GGTCTATAGC ACAGGTTCTA GGACAGCCAG GGCTACACAG
1141 AAAAACCTG CCTTGATTAA ACCAAACAA ACCAAACCAA ACCAAACCAA ACCAAACCAA
1201 ACCAAACCAA ACCAAACCAAG ACCAAACCA AACACTGAAG ATAGAACTTC AGTATTCCAT
1261 TCCTAGATAT ATACCCAATG GAGACTAAGT CAGCAAGACA CCTGCACAGC CATGTTCACT
1321 ACTACACTGT TCACCCACAGC CAGGCTGTGG ACCAGCCTG AGTGTCCATG ATAATGAAT
1381 GGATAGGTAA CTTTCAAGGT AAATGGACTC TGCTGTGTAC ATGCCTCACA TTCTGTTTAT
1441 TCATTTTCT TTATGAGGTG TCCATTCAAG AGTCACATGG TAGTTCTATT TTCAGTCTTC
1501 TGAAGATACT ACACGGTCC CCACAGTTA CACTTTATC AGCAGTGAAT AAGGGTTCCCT
1561 CTATCCTTAC CATCATTGT TGTAATTTC CTTGATGACC CTCTTCTGA CAGGGATAGG
1621 ATGTAATATC AGTGTGAGGA AGTACAACCT GTTTCTAAG TATTTATTGG CCCCTTGCAT
1681 TTCTTCTTTT GAAAATGTC GGTTCCCTGAC ATCTGCTCAG GTATTCAATTG GATGTTGTT

```

Figure 5 (continued):

1741 CTTTGGTGTG TGAGTTCTTA TGAATTCTAG ATGTTAAATC CCTGCCTGTG GTTCTCTCCC
 1801 ATTCTGTAAG CTGCCTCTC ACCCTGGCAA TTGTTGTCCT TGTTTGCAAG AAACCTTTGA
 1861 CTTCATGGAA TCTCATTGTC CAGTTTCCC TCCTCTGCTA TAGCCTGAGC TAATGCACTG
 1921 GTTTTACAG AGCCCTGGTC TATGCCCTTA TCCTCTCTG GCAGCTTCGG AGTTTCATT
 1981 CTTACATTG GATCTTTGAT CCACCTTGAA CAAGTTTGAG AGCAGGGTGA GAGATACGAA
 2041 TCTAGTCCA TTCTTCCATA TGTGATCTA GTTTACATAG CATCGTTGGT TGAAGAGGT
 2101 TTATTTATT TTTAATAAT GTGTCATAAA AAACGAGGTG GTTGTAGCAG TGTGGATTG
 2161 TTTCTTGTG CTTTGATCTA CAGGTCTTGT TTGTTGTCAG TCTCATGATG TTTTATTGCT
 2221 ATGGCTCTG CATACTGCT GAGGTCAAGGT ATTGTGATAT ACCTTCAGTA TTGCTCCCTC
 2281 AGACTCAGGT TTGCTTGGC CAGGAGTCAT CTTACTCAGT GCTCTTAGAG CTCCCCCAGC
 2341 ATGTAGCTG TACTATTCTT AGTTGATAAA TCAGGAAACT GGGGCTCAGA GAGATTAAC
 2401 GTCTTGAAC ACTTCTGGGG AGGTGAAACG TGGAGACACT AAACGTGTT TACCCGTAC
 2461 TGCTCCAGTA GCTGTCGGGT GCTGGGCTAC AGCAAAGCAC CTATACTATA TATTACTCAG
 2521 GAGGTGGAAA AACTCAGCCT CCCTTGGGGT TCCCAAGCTC CCAGGTGTCC AGTCACTGCT
 2581 GGAAACCTCA TGGAGTCTGA AAGGAAGGGT TGAGGGTACA TGGGGCAGCG ATGAGGAGCC
 2641 TGGGGCTGGG ATCTCCCAA CACCTGGATA TCCAGATGCC ACTGGGTCAG GGGGAGTTGG
 2701 GAACAGAGTT GGGATGTCGA TGGACCTGTG ACAAGGCCAG GGCCAGGGGG AGGATAACTC
 2761 TGGCTTACT AATTGCGAA AGTCCTTAGC TTAGCAGCAG TTGTCGGGA GCACAGAGGG
 2821 GCCTCTGTA AGAGGCTCAG GCAGTGGCC TCTGTAGGCG AAGGTCTCT CCATGTTCCC
 2881 CATGGTGGTT CTTGATGAAA GAGACAGTCC TTGGCTCAA ACTGGTTAT TGATTGTTCA
 2941 TTGTTGAAAAA TGGGTGCACA CCACCTCTC AGGGTGGACC AGAGATCAA TACCTTTGC
 3001 AGGGAGGAAT ATCTGGGAAG GGACGCTTAC TGGCTAAACC CTCAGGGCCT CTAGATACAT
 3061 CATTAGCATG GAGAACTCTG TTCTGGCTA CATGACCACA GGCCACATT CCACAAGCCA
 3121 CATGTGGAA GTGTGGCACA TGTCTAGGC CAGGAATCTG GTAGGGAGCG TGGAGCCACC
 3181 TACCATCCA GGTGGGTGCC TGGGTGCCAG GGACCTGAA CCCGCTCAAC CTTACCAAGT
 3241 TTCTGGCAG GGTCCACTGT CCTACACAGA AGCTGGAGGA GGTGTGAGGG TTGTCCTTT
 3301 GTGAATGTC CCATGCTGCT TGGGGCTCAG TTTCTCCACC TGTACCTCAT TGGTTGGGT
 3361 ATAAAAAGTG GGGATACTTT ATTATTCTCT GACTCGGTCC TGAGGAAAAA GCATCGTGGC
 3421 AGTCCAGGAA CCACACCTCG AGGTTCTGC ACTGAAGGGG CTCCCTAAGT CTCCTGGAGTC
 3481 TCTCCCTTC ACAGAGCTGC CAAAGCTAG GTTCTTTGA GGATAACAGA GCCATGCTTG
 3541 GTAAGCAGAC AACAGCATTG GTTACTCAA CCTCTTTTG TCAGCTCCCT CTCATAAAC
 3601 AAAGTGAGAC ACCATGCTGG CTTGAGGAAG ACTTCTAAAG CCAGACAAC GTGCAAGGAA
 3661 GAAGAAGAAG GGGCAAGTGG AGTTAGCTG GATGTAGCCC TCAAAGTCTC CAGAGACCAG
 3721 CCATGAAGGC TCAAGTGGAG GGCAGACCT GCAGCAGCCA AGCATCTGGC AGGAGAGGAT
 3781 CCTGGGAACC CCTCTACCAT GACACACATT CTTCTGCAG GTCACACTTA ATAGGCCATT
 3841 TCTTATTGATG ATCTATCATG GTGTTCTGTG CGAGATTAAT GAGGTGTTAT GCTGCGAAC
 3901 GAAAGTTATA TAAAAACAAG TCCCCCCCCC TTGTCACTGC TGCTAAGAAT GTAGCAGAAA
 3961 TTGTCCTCAAG TGTCTCTCTA ATCAGAAACA ATAAAGTCT CCTTGGATTC AAGCCCTCCA
 4021 GTTCTCTCT TCCCTGCTGA GCCTTGGACA CCCATACAAA CCTCTGGAT GCTACAGCTC
 4081 TGGGAGAGA CTCCAAGGTG GGGAGAGACT GATGGTACAA AAGCAAAATA CTTGTTGGG
 4141 GGTACACCCA CTCCCTGCC TGTGTTGTT CTGCACTGC TCCCTGCAGAC AGGCCCTCAG
 4201 TGGGTCTTCC ATGGGCAACA CGCAGAGGG GGCATGGAT GGGAAATACCC ACACCCCTGGT
 4261 TAGTTTACCC CGGCCATGCT CTCTGCTCT CATCCCTCTC CTGCCCTCTG CCACGGCTTT
 4321 CTCTGCAGGA ATCATATCTT CATATTGCC CACAGGTGTT CTCTCACCC TAGCTATGAT
 4381 GTTTACTTTA GAGTGACCTT ACCAGGGCTG GTGGGAATGA GTTCTAGAAG GCTCACGGAG
 4441 ATGCTAGGGA AGAAACGTCT TCTAACTACT GAGGTTACTA AGTTCTGGT GGTGTCCTCT
 4501 GCCTTCCCT TGTAAAGTC ACCTTGAAGT TAGTGCAGAA GAAATCAGAG CCCAGTCACA
 4561 GAGTAAATAT GGTCTGAAAG ATTCCTTTG AGTGCCAGA ATCCATGACA TTTCAAGAGC
 4621 CCTCTTGTA CCTTAAGTCA TTTGGGGTTG TATCTCTGC TTGATGTATG TGTGTCGTT
 4681 TATCAAAGAG TGAGATGGTT ACATAAGAGG TGCTCTAAAG GACAGAGAGG ATTTGCAATT
 4741 GTGGCATGTG ACATCCTCAG GCCTTGCCT GGTGCCAGGA GGAACGTGATG CAGAAAAGAG
 4801 TAAGAGGTCA TTCTCTGGAG GCTGTCATA TAGAGGAGAT CTTACAGTGC ATTCCCTCCT
 4861 CCAGGCCCTG CCTGAGGATA GACATGTGCT GACTGCAACT GAAACAGAGG CTTGGGATGG
 4921 AGAGTTAGGT TCACAGAAGG GAGGGTGGGA GATGGATGCT TGCTGGGTT TGGGTCCT
 4981 CACCAAGCTC TGACCAACCG GTCAGCCCAT GTGCTTATTG CATAGCTTTC TTTGCTATG
 5041 TTTACTCAGT GTGGTGTGTT TTGGGACCA GCAGAAGCCA GTCCCAGGCT GACAGCTGTG
 5101 GATACACAGG GCAGCATGAG GGTCTCTAGC CTGAAGCAGT CAGGCTGCCA GAAGAGAAAG
 5161 ACCAGCACAC ATTCTTCAA CCAACTATGT CTTGAAAAAC AAACATATTA TATCACATAT
 5221 ATTGCTTTA TGAGACAGCT AAAATGTACT CGGGTAGCAT GACTCCAGGT GGGGATATCT
 5281 GCAAGTGCCA TGAGTGGCAG AGGGACAGCC AATGTGAGGC AAGAAGGAAT TCTGGCTAA
 5341 CACAGCTTAG CTCCCTGGTG TTGGTTCAA CTTTGAGAGT TTGACCAAGA GCACCTTATT
 5401 TTTGACATAT TTAAACAGAG CACAACCTTG GGAAAAAGTT TTCTTATGAA AATTATCACA
 5461 ATAAAGCTTA AGGCATGACT ACATTAATAA GCTTTGCAA AGTATATGTG CCCTCTTCCA
 5521 CAAGAATGGT TCTATTGACT GAGAAATAA GTTCAGGATA AAGATCCAGG AAGAAGAGAT
 5581 CAGGGATAAG TAAAATACTA AACTCTTTG CAAAGTACAT AGACCCCTCTT TCATAACAAAT

Figure 5 (continued):

6441 GGGTTCTATT GACTGACAAG CACTGCTCAG GAGTTGGAA AGAGTCTAGC ATAAGCACGA
5701 TAGCCTGGAG ACTCTAGTGA GGTCTAGTCT TACAGACAGC AAAAATCACC AGGTTACAAA
5761 CTACATTCACT TTCCAGTTT CTGATCAGGC ACAGGTATGA ATCCCTTCTG TTGAAGAGAA
5821 AAGTCCATGT GTTTAAAATA TCTGGTTCT CCAGTGCTAT TAGGGAGAAG ACTTGAGCCC
5881 TATACAACTC CCACCTGGAG TGACATCCTG TCTTCATGGT ATATTACATA CCTAGACAGC
5941 CTCATCTCAC AGACTTAGGA CTTTGTCTTC TGATCTCCAT TTCTGATCCC ACTTCCACCT
5001 TTGCCCTGTAT AGTGTCTTT TCTTCACTG CTTGGTACCA ACCATGTTAT CCTCTGTGTA
5061 TTTGAGTGTGTT ACCATTTCAC GATTTTACCT GTATGCAAGA TCACACAGTC TTTGTCTTC
5121 TGCTCTGGATG CATGCTAAC TCTACACAAAC AACCCTTCCC CGTCACTCAG ATCTTCTCC
5181 ATTAACACAT ACATGGTCT GAAGAGGCTA GGGAGCTTCC CTTCACTGGG GAGCTAGCTG
6241 GCTATTGGGC CTTTTGACT GTCCAGGAAG GCCCCCAATT GCTGAGACAA GAACCTAGAT
6301 TCTTCATTAT TGACTCTAAC TCATGTATCA AGCAGAACG AATGAATAGT TATCAACAGG
6361 ATCAGGGTT CCAGTGTAAAG ACACATTGAC ATGAAAGAAC GGAGGAAGGA CAGATGGATG
6421 CATAAAAGCA GGACCACTGC CCCAGGAAGG TCCTGAAAC TGATGCAGGG CAAAGGACAG
6481 GTTATAAACCC AAATCTTAGG GAGTCAGGAA GAGCACAGAG GAGCTCAACC AACTGACAC
6541 TGCTTAGGGG CTACCAACCC AATCCTCCCT GTGGGAACAG CTAAGCTATC AGCCAAGGGT
6601 AATAAACAGG CAGGACCTGT GGATGACATG GAGAGCATAG GGACCCCTGGG TCCAGCCTT
6661 AGCACCTGCA CTCTCAGGAT ACTCCACCAT TGTGCTTAG AGAGCCTAGG GATACTGGGT
6721 CCAGCCTTTG GTACCTTCAC TCTCAGGGTA CCCCACATCACT GTGCTTGGG GAGCTTAGGC
6781 ACCCTGGGTC CAGCCTTCAG TACCTGCGCT CTCAGGACAC CCCACCATTTG TCTCTTGGCC
6841 CGTCTCTTCT TCCTCTTCTC CCCTTCTCATT GTCTCTTCTC TGTTCTTTC TTGACTCTCC
6901 TTCCCCCTCA CACCCCTCACT CTAGTTCTC CTTTCCCTCT CGTCATCACC CTATTCTCTC
6961 TGTGGTCCCT CCACTTCTCT TTATCTCTCA TGCTTCTCTC CTTCCCTCAAA TACTTGTCAAC
7021 CCACTAATCTC TCAGGGGCCA GCTCTAGTGA CAAAGCTGTT AATAGCAAGA CTCTCAGATC
7081 TCCAACGGCT CAGAGGAGCC AGACCCACCA AGAACTCTCT CCAGGTCAAATTTCAAGG
7141 CTTCGAAAGC TTTCAGCAA TGCTCAGGGAA ACATGCCACT ACAAGAAAGA TGCAAATTCC
7201 AGTTGAGAGT GGGAAAGGCC CTTGCGTAGG TCCCACCTTC CAGGCCAAGG TCAGAGGGC
7261 TCTGTGTAAT CCGGATTGAC AGGGCTCAGA ACAATGTTT GTTTTTAAGG TTTATTATT
7321 TTAGGTGTTA GTGCTTTGC TTGATGACC TTATGTGAT CATGTGTGTC CAGGGTCTG
7381 ATGACAGTAG AGGAGGGCTT TGAATCCCTG GGGATAGGAA GTACAGGAA ATTATAAGCT
7441 GCTTGTGGG TCTTCTAGCT TTCCCAACAG AAGTGAATGC TCTTCACCA TGACCATCT
7501 CTCTAGGCC AAGAGACATT GCTTTATGGA TATAATTGTG TGTTGTGTC AACATTGAGG
7561 AAAGGGAAAT AAAAAAACTT CTCAGGCC CTAAGGTGTA CAGTTTCACT AATTGCTACT
7621 TTATGTGTC ATAAAATGGC AGGTGCTTC ACATTTATAT ATACAAAAC TTCCCTGCTG
7681 GTGGTTCACT TGTGAGAACT GGGGTAAGTG GGTGAGTTCT CTTTTCTGT CTCTGCTCT
7741 GTCTCTCTC TTCCATTCTT TCTTAAAGGA AATAAACATT GCAGCTGGGT TATAGCTCAT
7801 CAATATGGAA GTTACAGAAAG TGAAAAAAGG CATTGCCCTG GTGGGTGGTG TTACCACTG
7861 ATTTTGGTT GTCTGCAAG GAGGTCTGGG GACTGGCTGC TCTGTCTCTG TCTGTATGAG
7921 TGAGGGAAAGT CTGGGGAGCA GATTCCTAA CCTTCAGCCT GGCCTGGTT CTGAGTGAAC
7981 CCAGCCTCTC TGGCTCTAGT AGCTTTTCC AAACAGGAAT CTGAGTGGTG ACAGGAAACA
8041 AGTACCAAGCC CATTGTTAA GTGCCAGGG TAGTGAGGC AGGAAGCTGC CATAGCTGGG
8101 ATTAGTAGTT GTATTGGATG TAGGAAGTCC TATCTGGGA CAGCTAAC TTAATGCTTC
8161 ACTGGAGATT TTCAATGAGA AATTATGCC ACGGCCATA TGGCCCATC CTTTTCTC
8221 CAACAGCCAA GTATTTCCTA TTAGAGGAGA CTTCTGTAC ACTTGATGGA TGTCATCATT
8281 AAGGTGACTT GGGGCACTCA GTACAGACTT GGGATGACCT CTGACAGCCT AACCTCTCCC
8341 CAACAAGGGC CCTCTATGTT TGCTATGTTA TGTAATGTCA GACATTGTCA GGAGTGTCCG
8401 CAGCAGCACCC TGCCCCAGTG GAGGGCTCTC ATAGGTTCC CACTGTCTTA TCTACACAGG
8461 GATAACAGGG AGGTAAGCTG CAGTTCCAG TCTCACTCA CAGAGGAAGA GATAACCCCA
8521 TCCCAAGGTCA TGTAGGCCAG AGTGGAAAGA ATGAGGATT TGACTCAGGT CTTCCAAGTC
8581 CCATTGATAG CATCTCTCA CAAGTCCCTT GCCACCCCTCA CGATGCCCTA GACACTTGCC
8641 TGCCCTTAT ACTAAGGAGA TGCAGGTACA AGGGGTTAC CCATGTAGCA GCTGAGGCG
8701 CTGGGGATAG ATACCAGCAG CAGGCCCTGAT GTCACCACTC TAACTCCAGC ATCCCCAGTC
8761 TGTGTTCTG GAGTGTGAAA ATCCCTACTT ACAAGATTG TGCAACAGTC CTGGCTCTG
8821 TGACCCATAG CTGGAAACAG GATTCTCATT GATTGTGGA ACATGGTGGC AGCCAGCCAA
8881 AAAGAGGGTC TGCATACAGA AGACACGTT GGCAAGGCCA CAGCAGACTC TGACTACCTT
8941 AGCTTACAGA ATTACAAGGT CATAATGTCC TCTGCTTGG TACCTCTCAT TAAAGGACAG
9001 GCCCTAATGA AGATGGGCA GAAGACTGAA GGAATGGCCA ACCAATAACT GGCCCAACTT
9061 GAGACCCATC CTACAGGCAA GCATCAATC CTGACACTAC TAATGATACT CTGTTATGCT
9121 TGCAGACAGA AGCCTAGCAT AACTATCCTC CGAGAGGTCC ACCCAGCAAC TGACTGAAAC
9181 AGAAAAAAGAT ATCCACAGGC AAACAGTGGA TGGAGGTCA GGAAGTATTAT GGGAGAGCTG
9241 TGGGAGGAT TAAAACCCCT GAAGGGATA GGAACCCAC AGGAAGACCA ACAGAGTC
9301 CTAAGAGACC TGTGGGAGCT CTCAGAGACT GAGCCACCAA CAAAGAGCA TACACAGGCC
9361 GGTCCGAGGC ACCTGGCAGC TGTGAAGCAG ACATGCGACT CAGTCTCCAT GTAGGTCTC
9421 CAATAAGCGG TAGCCTGACT GCAGTATCCA ATCCCCAACA GGGCTGCATA GTCTGGCCTC
9481 ACTGGGGGAG GATGCCCTA ATCCTGCAGA GACTTGATGA GTGGAGAGCT ATCCAGGGGG

Figure 5 (continued):

9541 AACCCACCCCT CTCTGAGAAG GGAATGGGG A TGGGGGAGGG ACTCTGTGAA GAGGGGACAA
 9601 GGACAAACAA GAACCTAAA TAGGTCAGGC CCTAAAGGCT TGCTAAGTAG CAGTGGCCCA
 9661 GCTCTGTCT GTTCCTCAGC CCAAGGCTCA GCTCCACCT GTTCTGTGT TTTCTGGCT
 9721 TTTCATGGC CTAGGACTTG GTGACCAGTT CAAACATGG GGCCTGTGGA AGACACAATA
 9781 TACAAGACTA GGGACATTCC TGTCTGTG ACTATCCATA GCCTGATGTA GGTGGAAGGA
 9841 CCCAACACTA GGATTCTAC CTTGCAACAA CCTTGACAGC TGAGGGCCTC TCAGAACCT
 9901 ATTCCTTCA CTGAAAAATG AGACTCTCAA ATGAACGTC TGACAATCAT CAGGCTTATT
 9961 AAAGAGGTT ATCTAACCTG ATGGCAAGC AGACAGCAGG CAAATGTCG TATCAACCTC
 10021 TAGGAAGGAC AAGAACTGCT CACTGCTGCC CCCAGGAGG CCATTTGCTG AAACAGCTGC
 10081 TCTCTGTG TGACAGCAGGC CCTGCCTTCT CATTGAGGCC ACAGCCCCTT CCTGTGTGAA
 10141 CCTCTGTCA GGTCACTGGG AAACAGATCA AGATGGAACA GGACAGCTCC TGATGGTAA
 10201 TAAAAAACAG TGGTCATGGC TATTCTAGG GTTTATGCT TCTTCAGTCC ACAGTGTGAA
 10261 GAGCTGTGGG CATGAACAC AGTGGTCGAG GTAGAGTTGG GTTCTGAAA TTCAACAGTGG
 10321 GGTGAGCTCA GTAAATGTGA GCTGGAGGTC ACTCGTGAGA CACACAGTCC TGCTGTTCT
 10381 GTTCCAATA TCCTGAGGAG ACGACACATC TACTTGTTC AGAGGCCACA GTCTAGTTGA
 10441 CCTGAGAGTT ACCAGTTCT TATTGTTG TGTTGTTG TGTTGTTG TGTTGTTG
 10501 TGTTGTTG TGTTGAGTGC AGGTGCACAT ATGATAGCGT ACACGTTGAG GTCAGAGGAT
 10561 AACTATCAGG CGTTGTCCTC TCTACTTTT CCTCGGACTC TGGAGAACAA ACATGGGTCC
 10621 TTATTCCAGG GGAGCAAGTC GCTGGTGGT GACACATCTT GTCACATAC ATTTTACCTA
 10681 GACAATGGAG CCTCCATCAG AGTATTACT TAGCTCCTCA CCGATGGCAA TGCACACCT
 10741 CTCTACCCAC ATAGGAGTTG GGTCTCCACA CACCCCCACA CCCCCCTTCAC CAAAACGTTT
 10801 TCAGTTACTT TATCTGGTAA AGTTCATCAG AGAATGAAAGC CAGTATTAAG AACATGGAAT
 10861 CATTGGGAA CCTGGATCTA GCAATACCCC ACCCTAGATG GAGTTGCTG GTTTTACCT
 10921 CAGATTATAA TTCCCCCTA GCTTCTATGG TTTATTCTGA ACCCAGGGGA ACTCGATTCC
 10981 TCCCTTGGGAA CCACAGACAT CCTGGCTTGT GAATTACAT GTCATCTACT GCTAATCCAT
 11041 TGGTAGTATG TGGCTCACAG AGACACACTA CAGTCATGGC CAATGTCAG GTAGGACAGA
 11101 TGTGAATCAT TCCCCCAGTC CTGCTGTTT CATGACTAAC CCTCCTCAGC ACAGTGACCA
 11161 TGAACCTACT TTTCCCTCC TTTTATTTT AGAATTGCTG GAATTCTA TTTTGAGAAA
 11221 TAATAGCCTT GGGCAGCATT AAACAAAATC ATCTAGAAAAG CTGGTTAAA ATACAGATGG
 11281 TTGAGTCAGT GAAAGAGTGA GGAATGTCAT TATTGGCCCCC TCACAGAGGC TGGCTCACTC
 11341 CAGCAGAGGT GGTGAGGCT CTTGGACAGC GGTCAAGGTGC ATAGGAAAGG TNGTCTGGGA
 11401 CACTGAGAAC CACAATTGAA CAAACAGAAC TGTGGCTTT TTTTTTTTA AATGAGTTCT
 11461 CAAAAAATGA CTGGCTAGCT TAGGCAAATA CTTGGAGCCA ACCCAACAGA ACATTCTCC
 11521 ATTGATTCTAT TCTGGATCTT CTTCTAGAC AATACTGAAC TGACCCCTTG TTGGCAGTCT
 11581 CAAGTTGAC AACATAGGGC TTTGAACCTG GCACAAGGTC CATCACTGTC ACCCAAGCAT
 11641 CCTGGGTGAC CTTGGGTTG GAATATCTG GCTAACCTTA GATATTTCT TTGGAGTATC
 11701 TTTAGAACAT CCAGGAAATA GGGCTTGATT CTCATCCTGG GACCACAATA TAAGTCACCC
 11761 TAGAATCCCA GGAGATCGTG CAGAGAAACA AGGATCTCTC TCGTGTGCA CTTCTTCAA
 11821 AGCAGTGAATG AGTGAATCCA CTAAACTGAG TTCCCATCTG AGAGTCCACA GGAGGCTTTG
 11881 GGGCAAGAAC CAGAGGGAAAG GCACTGTTG TGTTGGTAAA GTTTTGACTC TAACAAATT
 11941 GAAGACATAG ATGACATTGT GTCAGACTAA CAACAACTTA GACTCATGTG GTTCTGTTT
 12001 AGGGATCAGA TTTTATTCTAT CAATGACTTG TCTTAGTGTAG TAGAGAAAGG CTTCTACTG
 12061 GAGTGTAGGC TCAATAATGA CAGAAGAGAT AGCTATTCTC CCTAGGGACT GTGCTGCTCC
 12121 AAGTTGGTG GAGAAAGGC GTGGGGAAAC TAGATGTCCT CTCCTGGGAG GGGGCTCTGAA
 12181 GCTGGCTTCA TAGAAGGTGT GAAGTTTGC TGAAACATCT AAACAGAAATT ATAGCTTAGG
 12241 AAAGTGAACCA GGCAGGGCAG GGAATGTGTT GCATATGTAT ATGTACATGA ATATATTATG
 12301 TTATAGATAC ACACACATTG GAACCTCATT TGCAGATGAC AGAAAATAGG TTATTTGCC
 12361 TCTCTTAATC GCTAAGCACA ATGACTTCA GTTCCATCCA TTTCTGAAA TGCCACAATT
 12421 TCATTTTCA TTGTTGGCTGA ATAAAATTCC ATTGAGACT GGGCCCTACT TCATCCACTC
 12481 CTGAGGGCAG GCATATCCCC TGGCTCATT TCTTACCTAT TGTGAAGAGA ACTGCAACTG
 12541 TCTTGTGAA AGGCAAGCGT GAGAGAGGCC GGCACATAATT GTGGGTTTT GTTTCTTCTT
 12601 CCTGCTATGA CTCTCCATT GTCAAGAACCA AAGATGATA AAAGCCGCCA CCATGAAAGC
 12661 CATCTTAATC CCATTCTTCT CTCCTCTGAT TCCGTTAACCC CGCAATCTG CATTGCTCA
 12721 GAGTGAAGCC GAGCTGAAGC TGAAAGTGT GGTGATTGTC AGTCGTATG GTGCGTGC
 12781 TCCAACCAAG GCCACGCAAC TGATGAGGAA TGTCACCCCA GACGCATGGC CAACCTGGCC
 12841 GGTAAAATG GTTGGCTGA CACCGCGCGG TGTTGAGCTA ATCGCCTATC TCGGACATTA
 12901 CCAACGCCAG CGTCTGGTAG CGCACGGATT GCTGGCGAAA AAGGGCTGCC CGCAGTCTGG
 12961 TCAGGTGCGC ATTATTGCTG ATGTCGACGA GCGTACCGT AAAACAGGGC AAGCCTCGC
 13021 CGCCGGCTG GCACCTGACT GTGCAATAAC CGTACATACC CAGGCAGATA CGTCCAGTCC
 13081 CGATCCGTTA TTTAATCCTC TAAAAACTGG CGTTGCCAA CTGGATAACG CGAACGTGAC
 13141 TGACGCGATC CTCAGCAGGG CAGGAGGGTC AATTGCTGAC TTTACCGGGC ATCGGCAAAC
 13201 GGCCTTTCGC GAACTGGAAC GGGTCTTAA TTTTCCGCAA TCAAACTTGT GCCTTAAACG
 13261 TGAGAACAG GACGAAAGCT GTTCATTAAC GCAGGCATTA CCATCGGAAC TCAAGGTGAG
 13321 CGCCGACAAT GTCTCATTAA CGGGTGGCGT AAGCTCGCA TCAATGCTGAA CGGAGATATT
 13381 TCTCCTGCAA CAAGCACAGG GAATGCCGGA GCCGGGGTGG GGAAGGATCA CCGATTCA

Figure 5 (continued):

13441 CCAGTGGAAC ACCTTGCTAA GTTTGCATAA CGCGCAATT TATTTGCTAC AACGCACGCC
 13501 AGAGGTTGCC CGCAGCCGCG CCACCCCGTT ATTAGATTTG ATCAAGACAG CGTTGACGCC
 13561 CCATCCACCG CAAAACAGG CGTATGGGT GACATTACCC ACTTCAGTGC TGTTTATCGC
 13621 CGGACACGAT ACTAATCTGG CAAATCTGG CGGCGCACTG GAGCTCAACT GGACGCTTCC
 13681 CGGTCAAGCG GATAACACGC CGCCAGGTGG TGAACTGGTG TTTGAACGCT GGCGTCGGCT
 13741 AAGCGATAAC AGCCAGTGGA TTCAGGTTTC GCTGGTCTTC CAGACTTTAC AGCAGATGCG
 13801 TGATAAAACG CCGCTGTCA TAAATACGCC GCCCCGAGAG GTGAAACTGA CCCTGGCAGG
 13861 ATGTGAAGAG CGAAATGCGC AGGGCATGTG TCGTGGCA GTGTTTACGC AAATCGTGA
 13921 TGAAGCACCG ATACCCGCTT GCAGTTTGTG AGGTACCCGG GGATCACAAAC TTGCCCTCTG
 13981 AAGAGGAAGA ACAGAAGGAT GCCACAACTC TCCTGCTGGC TACTCTCCAG TGGTTTCATC
 14041 TTACTTCTGA TGGCATTCC CTCTAGAAAAG TGCTACTATC ATCCACACAT TTCTACCTGA
 14101 GACCACCCAA AGGACCCCTCC CAAATTCTCT TCCTCTCTGA GTAGTCTCCA CACCTGTTAC
 14161 CACCATCCCA GAATTTAAAT CTTAACTGCA CTCTGGCGTG TGACTTGCCT CAGTCCTTGC
 14221 AATAAGAGTT GTTGCAGTG CGAGGCGTGG TGGCGCACGC CTTAAATTCC AGCACTTGGG
 14281 AGGCAGAGGC AGGGCGGATT CTGAGTTCGA GGCAGCCTG GTCTACAGAG TGAGTCCAG
 14341 GACAGGCCAG GCTATACAGA GAAACCTGTG GTCGAAAAAC CAAAAAAA AAAAAAGTT
 14401 GTTGGCAGAG TGTGGTTAT ATACAGGTG GAGATTCAA ATGAGTGGCT GAAGCTGTAG
 14461 CCAGAAGGAA CTTAGAGGT AGCTCATAAC TAAAAAGAA ATGTAGAGAG TAGCAGAAC
 14521 ATTGAGAGAG TGGGCACACA GCCACTGTGT GAATGTGGCA GAACACAATC CAGCCAGCTA
 14581 TACATGCATA AGTGTATATT GGCGCCATCC TGACTGTATGA GACACAGGAA AACAGATAGA
 14641 CGGGGTTAGG TGGCCATGGC CTTTCTGCC TGCCCTTCC TAAGGGTCAT CTCAAGACCT
 14701 TATGCTCTCT TAACTCTTCC ATTGCTACTT AGCTTCTAGA TATCACCTCC AGATTAGTCT
 14761 CCTTGGGTAC ATCAGTGTAC CTGGTGATAT CCAGGGCTTC CTGATTCCAT CTTTGTCTA
 14821 GAGGCTGCAA CAAAGAGGT CTTCTTAATA CTTCACACCC TGATGCCAAA AGGAAGACAC
 14881 AGAAGTTCAC AGAGGTGAAG TGATTCTATG AGGACATACA GTGAGCAAGC ATCAGGGTCC
 14941 GGATTATCTG ACTCTACTCT AACTTTATG TAAATGTGCT TTATGCCATT AACACTGTCA
 15001 TTCCGTGCT TCAGCTCTGG GAGACTCCA AGCCTCTTA GGCACAAGCC ACAATTAAGG
 15061 GACTCTGACA CTCTGCATTC ATTAATTAGC ATGGTGGCT CTATGTTCC AGATTATG
 15121 TTGTTTCACT TTCCATATAG GCTATGAAGG GTGTGAGGAA ATTTTTTGGG GACAGAATTG
 15181 GAGGCAATCC ACCTCTCTCA GGAAGCCTCT ATCTGGAAAA GCTTACAAC CAGGGACAGT
 15241 AACTGTAGGC CCAGTCTTG GTGTCCAAA TGGGTTTAT GTTTGAATC TGCAAAGCCT
 15301 TCCATGTGCT CAAAGGTTTG AACATGGAGC CTCCCTCTGG TAACACTGTG TTGGAGGCTT
 15361 TTGAGACTGG ATGCTCTTGTG GTCCCATGTT TTGCTACATC ATCTGTCAAG ATATGACCA
 15421 GGCATGCTAC CAGCTACCAC AGACTATGCC TCTCAGCTT TCATGTTCTC CCCACCATGA
 15481 TAGACTTGTG TCTCCTAAAA ATGGAATCAA AGCAAACCTT TCCTGCATTA AGTTTTTTT
 15541 TTTCTGTTAA GTGTTGGTC ACAGGGACAA GAAAACACTC AATACAGATA ATTAGTACCA
 15601 GAGTTGAGGT TCATTGCTCT AGCAAGTGG ATCAAATTTC TAGGGCTTTG GAACTGATT
 15661 ATAAGAGACA TGTAGAAGAG TCTGAAGCTG TGGGCTACAG AGTGTCACTC AGTTTTAAG
 15721 AATAGTTAA TACACCATGG GAATTGTGAA AATCAGAATG CTCACACAAA GGCAGACAGG
 15781 AAAACGTGAG CATGTGGCGT GTGAGAGGGC ATAAGAAGGA ACCTAGGGGG AAATGAGCTA
 15841 GAAGCCATTC GGCTACGTTA GGGAACGTGT GTGGCTGTGC TTGGCCCATG CCCTGGCAAT
 15901 CTGAATGAGG CCAAATTTC AAGGAGTGG A CAAACTCGAT TGTCAGAGGAA AATATCAAGA
 15961 CAGACCAACCA CTCAGGCTAT GCGGTGTTG TGACCGACCA GCTACTCTTA GCCAGCTCTA
 16021 TTGTGAAATT CCAGAGCAAT TATCAGAGCA TGAAGATACA TACAGTTAG TGAAGTAAGG
 16081 GGTGTGGTC CCTAAGTGGA TGGTGCATAA ATCTATGTAG GTGATGCCA AGTGCACATT
 16141 GATAATCCAA AATATCAGCA ATGTGGATG TCTTCCAAGG AGACCTGTAG ACACACATT
 16201 TAGAACTTTG CTCATGGCTG TAATAAATAG CTAGCTAGAA ATCAATTCTC GAAGAGGTTA
 16261 GTCTGAGTTA CGGTTCCAGG GCAAACATTC AGTGTGGCA AGGAAGGCAT TGCAGTCAGG
 16321 AGCCAAAGGT CAGCTGGTCA CATTGCATCA AGAGTAGAGA GTCAAGGTGT GAGTAGAAAG
 16381 AGGATACAGG TTATAAAACC TCACTGTCCA CTCTCAGCAA TCCATTCTC CCTAAAAGGC
 16441 TTTACCTCT AAAGATTAA GTCTTCAAA CCAGTACCG TAGCCTGGGA ACAAAAGTTG
 16501 AAACAAATGA GCCTTGTGG GGCATTTCAC ACTTAAACCA GGGCATCACC TAGGAGGAGC
 16561 CCTGTGTGCA GTAGGAAGTG TGGCCTCTGT GTCAGGAATG CTCAGGCTAA TAAGGGTCC
 16621 TCTATCTGAG GGACCCCTATG AAGATTCAAC AAGTAGTTGT GAGAATTCCC TGTAATGGA
 16681 TGCTACCAAT TTGACATTG TAGACCTGCT ATTGTGTGCT TCTTATTGG GCTCTCCCAT
 16741 CTCCCAACTT TCCAACCCAT ATTCCACATT AATCCCTTCC ACCACCATGC AACACTAGGT
 16801 AGGAGAGAAAG GAAGGTTAGA AGAGAAAGTG GGTATAGATC TATTTAGACT ACTTCCTGCT
 16861 GATTAGGGC AAGTCCAATC GTCAATTGCA GGATACCTCC AACCAAGCAAC CAGCAAACCA
 16921 GCAAATCAGA AACAGAAAAA GCAGCCAACA AGGCAGCACT AACCAAGCAGG ATTGGGTGCG
 16981 GTAGCGTGGG AGCAGTCACT ACTGGTCTTC TCATGGCTTT GGCATTAATA CTCTCTCAAG
 17041 AAATTCCGTA ATTGTTTCCCG CACCACCTGA AATTCCGTA TTTTAAATGC AAACATATCTA
 17101 CAGCTGGCAA AAATCACATC TCTCTAGAG CACAAGACAA ATCATAGTTA CTGGCTATT
 17161 GCAATCTGAA GCATCTCAAT ATCCACACC TGGGATTAAA AAAAAACAT ATTCAACATCA
 17221 CATAACTGTT TTTTTTCCCA AATTGTTAT TAGGTTTTT CTTATTAC ATTTCAAATG
 17281 CTATCCCCGAA AGTCCCCCTAT ACCCTCCAC CTCCTGCTC CCCTACACAC CCACCTCCAC

Figure 5 (continued):

17341 TTTTGACCC TGGAGTTCCC CGGTACTGGG GCATATAAAG TTGCAAGAC CAAGGGCCT
 17401 CTCTCCCAG TGATGGCCGA CTAAGCCATC TTCTGCTACA TATGCAGATA GAGACACGAG
 17461 CTCTGGGGT ACTAGTTAGT TCATATTGTT GTTCCACCTA TAGGGTCGCA GACCCCTTC
 17521 GCTCCTTGGG TACTTTGCT AGCTCCTCCA CTGGGGCTC TGTGTTTAT CTAATAGATG
 17581 ACTGTGAGCA TCCACTTCTG TATTTGACAG GCACTGGCCT AGCGTCACAT GAGCCAGCTA
 17641 TATCAGGGTC CTTTCAGCAA AACCTTGCTG GCATGTGCAA TAGTGTCTGC GTTGGTGGT
 17701 TGATTATGGG ATGGATCCAC TAGTTCTAGA GCGGCCGCCA CGCGGGTGGA GCTCCAGCTT
 17761 TTGTTCCCTT TAGTGAGGGT TAATTGCGCG CTTGGCGTAA TCATGGTCAT AGCTGTTCC
 17821 TGTGTGAAAT TGTTATCCGC TCACAATTCC ACACAACATA CGAGCCGGAA GCATAAAAGTG
 17881 TAAAGCCTGG GGTGCCTAAT GAGTGAGCTA ACTCACATTA ATTGCGTTGC GCTCACTGCC
 17941 CGCTTCCAG TCGGGAAACC TGTCGTGCCA GTCGATTAA TGAATCGGCC AACGCGCGG
 18001 GAGAGGGCGGT TTGCGTATTG GGCCTCTTC CGCTTCTCG CTCACTGACT CGCTGCCTC
 18061 GGTGCTTCGG CTGCGGCGAG CGGTATCAGC TCACTCAAAG GCGGTAATAC GGTTATCCAC
 18121 AGAATCAGGG GATAACGCG GAAAGAACAT GTGAGCAAAA GGCAGCAAA AGGCCAGGAA
 18181 CGTAAAAAG GCGCGTTGC TGCGTTTCC CCGTACGGTCT CGCCCCCTG ACGAGCATCA
 18241 CAAAATCGA CGCTCAAGTC AGAGGTGGCG AAACCCGACA GGACTATAAA GATACCGAGC
 18301 GTTCCCCCTT GGAAGCTCCC TCGTGCCTC TCCTGTTCCG ACCCTGCCGC TTACCGGATA
 18361 CCTGTCCGCC TTTCTCCCTT CGGGAAGCGT GCGCTTTCT CATACTCAC GCTGTAGGTA
 18421 TCTCAGTTCG GTGTAGTCG TTGCGCTCAA GCTGGGTGT GTGACAGCAAC CCCCCGTTCA
 18481 GCGCGACCGC TGCGCTTAT CCGGTAACTA TCGTCTTGAG TCCAACCCGG TAAGACACGA
 18541 CTTATCGCCA CTGGCAGCAG CCACTGGTAA CAGGATTAGC AGAGCGAGGT ATGTTAGGCG
 18601 TGCTACAGAG TTCTGAACT GGTGGCTAA CTACGGCTAC ACTAGAAGGA CAGTATTGG
 18661 TATCTGCGCT CTGCTGAAGC CAGTTACCTT CGGAAAAAGA GTTGGTAGCT CTTGATCCGG
 18721 CAAACAAACC ACCGCTGGTA GCGGTGGTTT TTTTGTGAG AAGCAGCAGA TTACGCGCAG
 18781 AAAAAAAGGA TCTCAAGAAG ATCCCTTGAT CTTTTCTACG GGGCTGTACG CTCAGTGGAA
 18841 CGAAAAACTCA CGTTAAGGGG TTTTGGTCAT GAGATTATCA AAAAGGATCT TCACCTAGAT
 18901 CCTTTAAAT TAAAAATGAA GTTTAAATC AATCTAAAGT ATATATGAGT AAACCTGGTC
 18961 TGACAGTTAC CAATGCTTAA TCACTGAGGC ACCTATCTCA GCGATCTGTC TATTTGGTTC
 19021 ATCCATAGTT GCCTGACTCC CGCTCGTGT AATAACTACG ATACGGGAGG GCTTACCATC
 19081 TGGCCCCAGT GCTGCAATGA TACCGCGAGA CCCACGCTCA CGGGCTCCAG ATTTATCAGC
 19141 AATAAACCG AGCTCCCTCG GTCTCCGAT CGTTGTCAGA AGTAAGTTGG CCGCAGTGT
 19201 CATCCAGTCT ATTAATTGTT GCGGGGAACC TAGAGTAAGT ACTTCGCCAG TTAATAGTT
 19261 GCGCAACGTT GTTGCCTATTG CTACAGGCAT CGTGGGTGTCA CGCTCGTCGT TTGGTATGGC
 19321 TTCATTGAGC TCCGGTCCC AACGATCAAG GCGAGTTACA TGATCCCCA TGTTGTGAA
 19381 AAAAGGGTT AGCTCCCTCG GTCTCCGAT CGTTGTCAGA AGTAAGTTGG CCGCAGTGT
 19441 ATCACTCATG GTTATGGCAG CACTGCATAA TTCTCTACT GTCATGCCAT CCGTAAGATG
 19501 CTTTCTGTG ACTGGTGAGT ACTCAACCAA GTCATTCTGA GAATAGTGT TGCGGGGACC
 19561 GACTTGCTCT TGCCCCCGCT CAATACGGGA TAATACCGCG CCACATAGCA GAACTTTAAA
 19621 AGTGCTCATC ATTGGAAAAC GTTCTTCGGG GCGAAAACCTC TCAAGGATCT TACCGCTGTT
 19681 GAGATCCAGT TCGATGTAAC CCACTCGTC ACCAACTGA TCTTCAGCAT CTTTTACTTT
 19741 CACCAGCGTT TCTGGGTGAG CAAACACAGG AAGGAAAAT GCGCAAAAA AGGGAAATAAG
 19801 GCGCACACGG AAATGTTGAA TACTCATACT TTCTTTTCAATATTAA GAAGCATTAA
 19861 TCAGGGTTAT TGTCTCATGA GCGGATACAT ATTTGAATGT ATTTAGAAAA ATAAACAAAT
 19921 AGGGGTTCCG CGCACATTTCC CGCGAAAAGT GCCACCTAAA TTGTAAGCGT TAATATTG
 19981 TTAAAATTGCG CGTTAAATTG TTGTTAAATC AGTCATTTT TTAACCAATA GGCGAAATC
 20041 GGCAAAATCC CTTATAAAATC AAAAGAATAG ACCGAGATAG GGTGAGTGT TGTTCCAGTT
 20101 TGGAAACAAGA GTCCACTATT AAAGAACGTG GACTCCAACG TCAAAGGGCG AAAACCGTC
 20161 TATCAGGGCG ATGGCCCACG ACGTGAACCA TCACCCCTAAT CAAGTTTTT GGGGTGAGG
 20221 TGCGTAAAG CACTAAATCG GAACCCCTAA GGGAGCCCC GATTTAGAGC TTGACGGGG
 20281 AAGCCGGCGA ACCTGCGAG AAACGAAGGG AAGAAAGCGA AAGGAGCGGG CGCTAGGGCG
 20341 CTGGCAAGTG TAGCGGTACG GCTGCGCGTA ACCACACAC CGCCCGCGCT TAATGCCCG
 20401 CTACAGGGCG CGTCCCATTG GCCATTGAGG CTGCGCAACT GTTGGGAAGG GCGATCGGTG
 20461 CGGGCCTCTT CGCTATTACG CCAGCTGGCG AAAGGGGAT GTGCTGCAAG GCGATTAAGT
 20521 TGGGTAACGC CAGGGTTTC CCAGTCACGA CGTTGTTAAA CGACGGCCAG TGAGCGCG
 20581 TAATACGACT CACTATAGGG CGAATTGGGT ACCGGCCCCC CCC

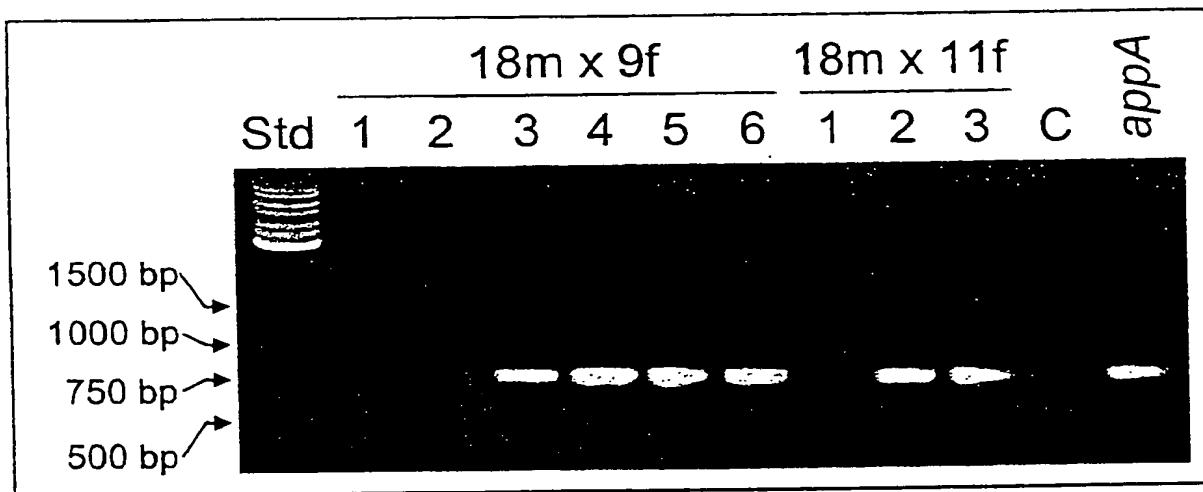


Figure 6

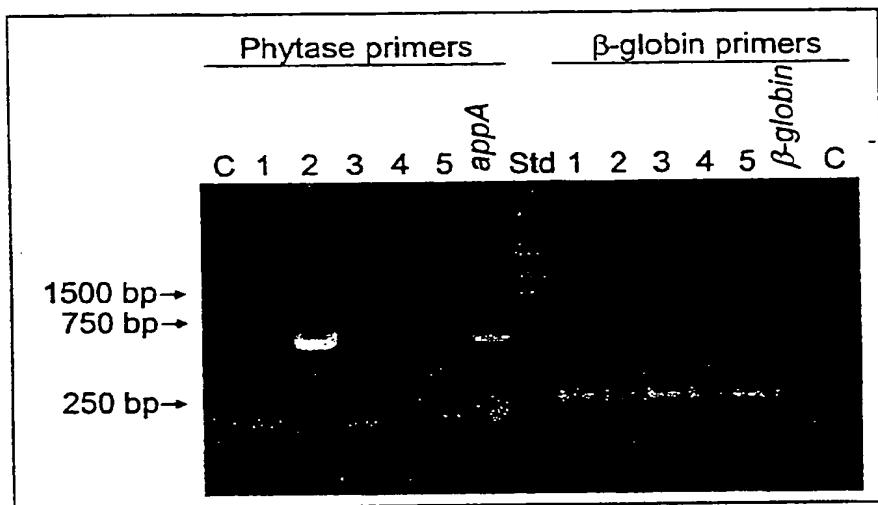


Figure 7

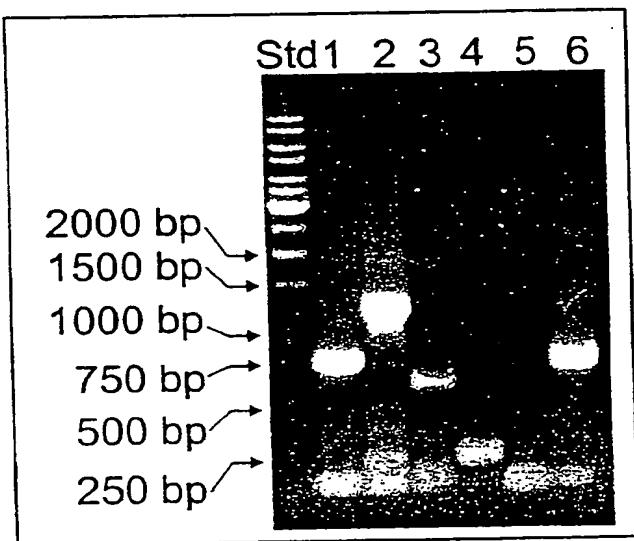


Figure 8

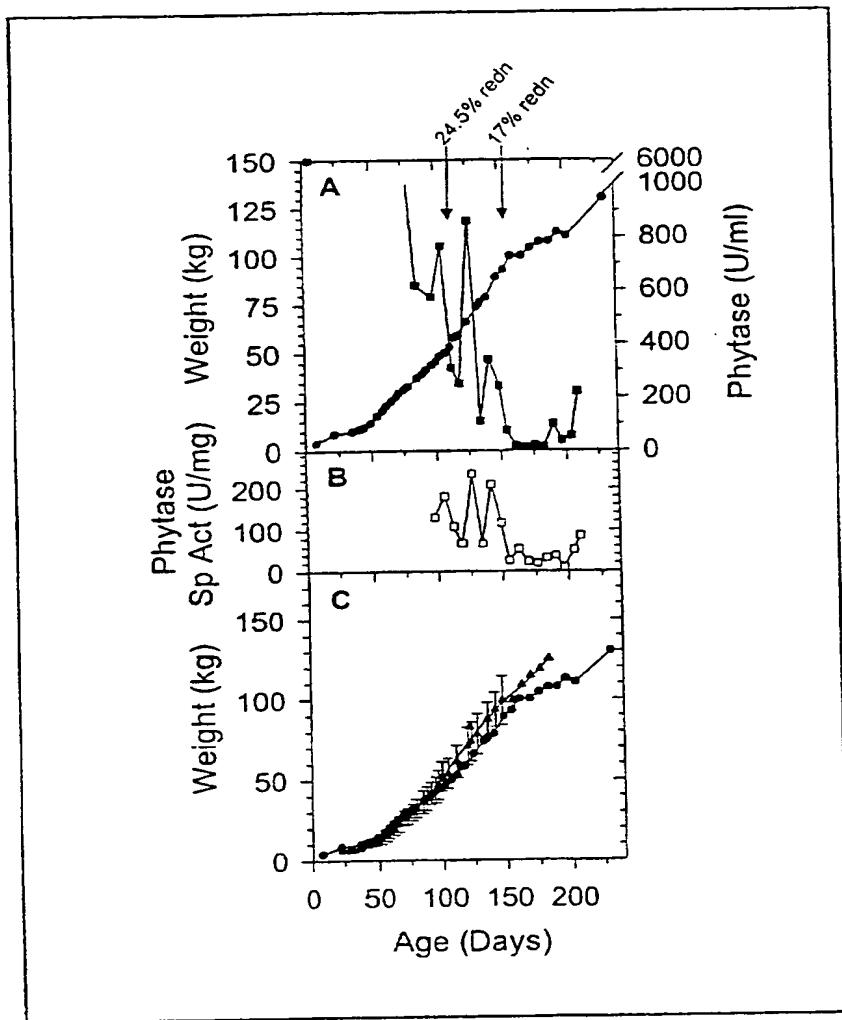


Figure 9

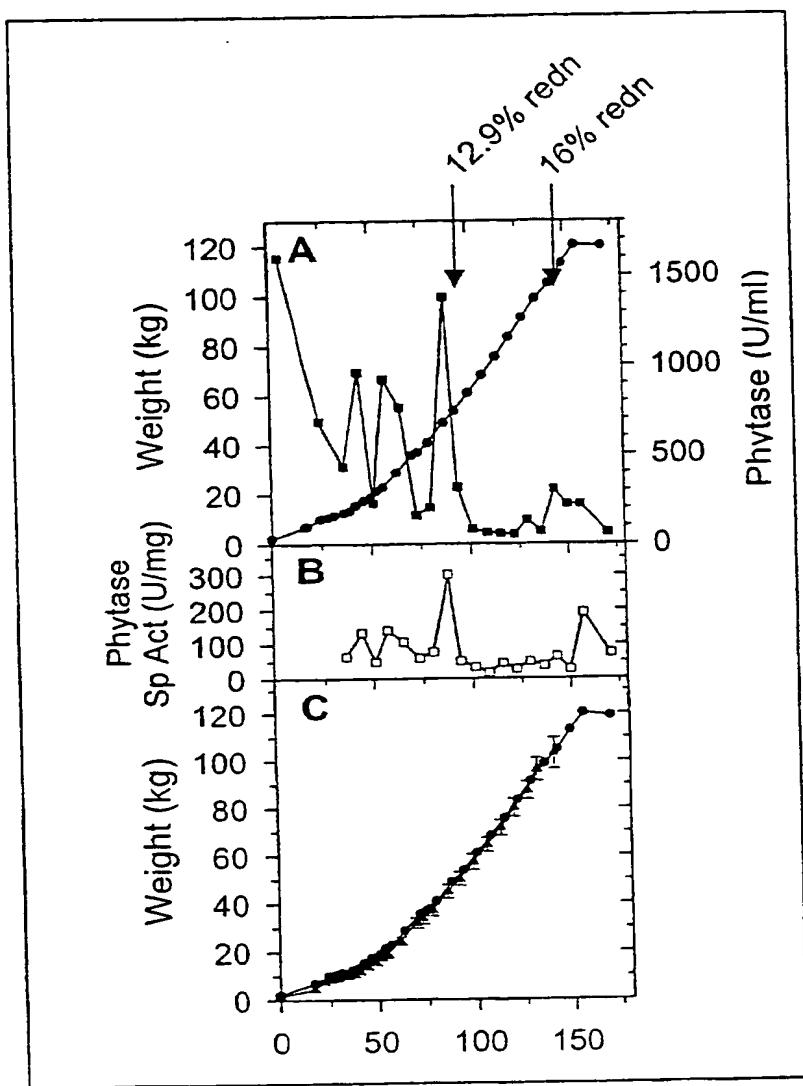


Figure 10

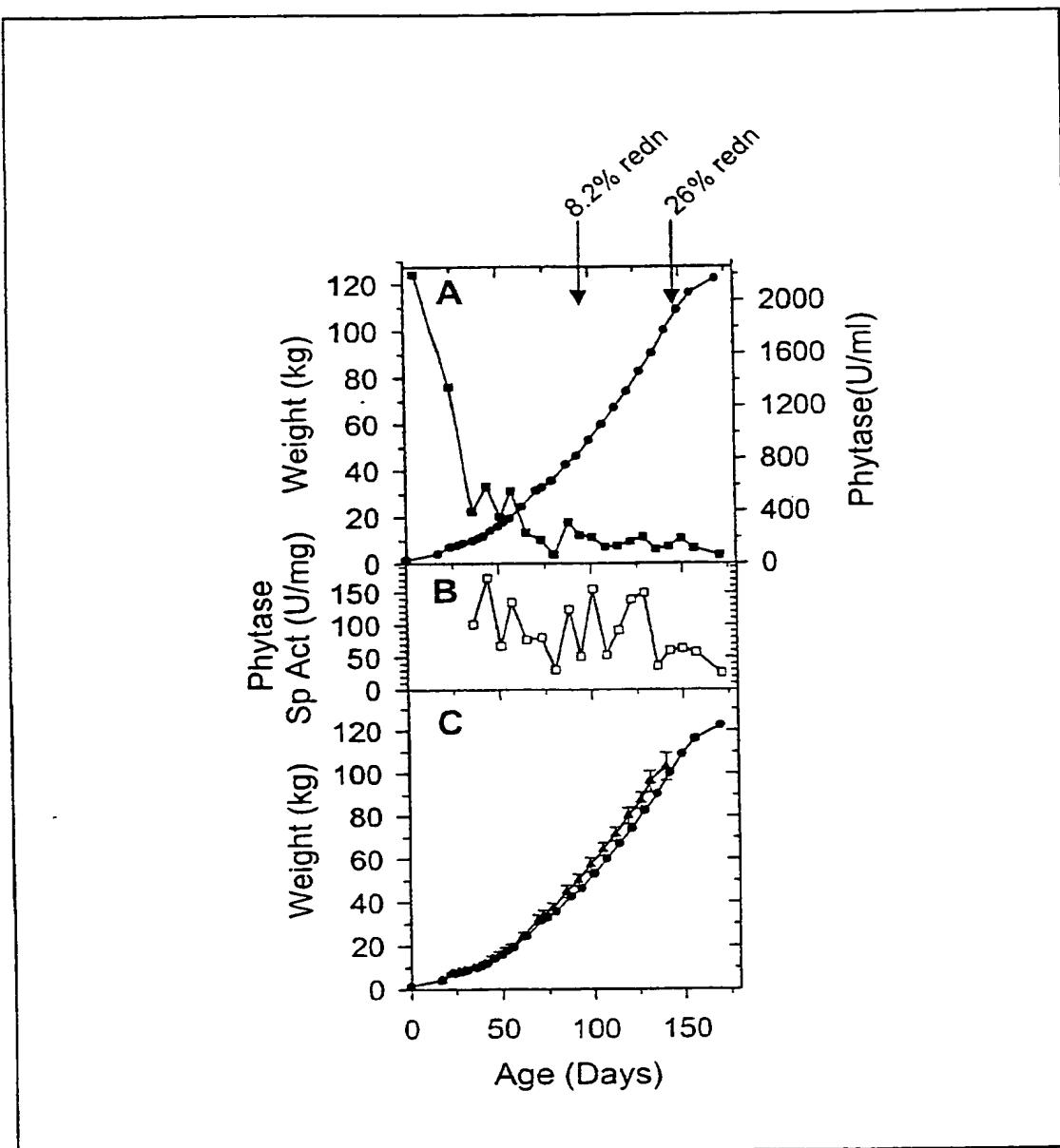
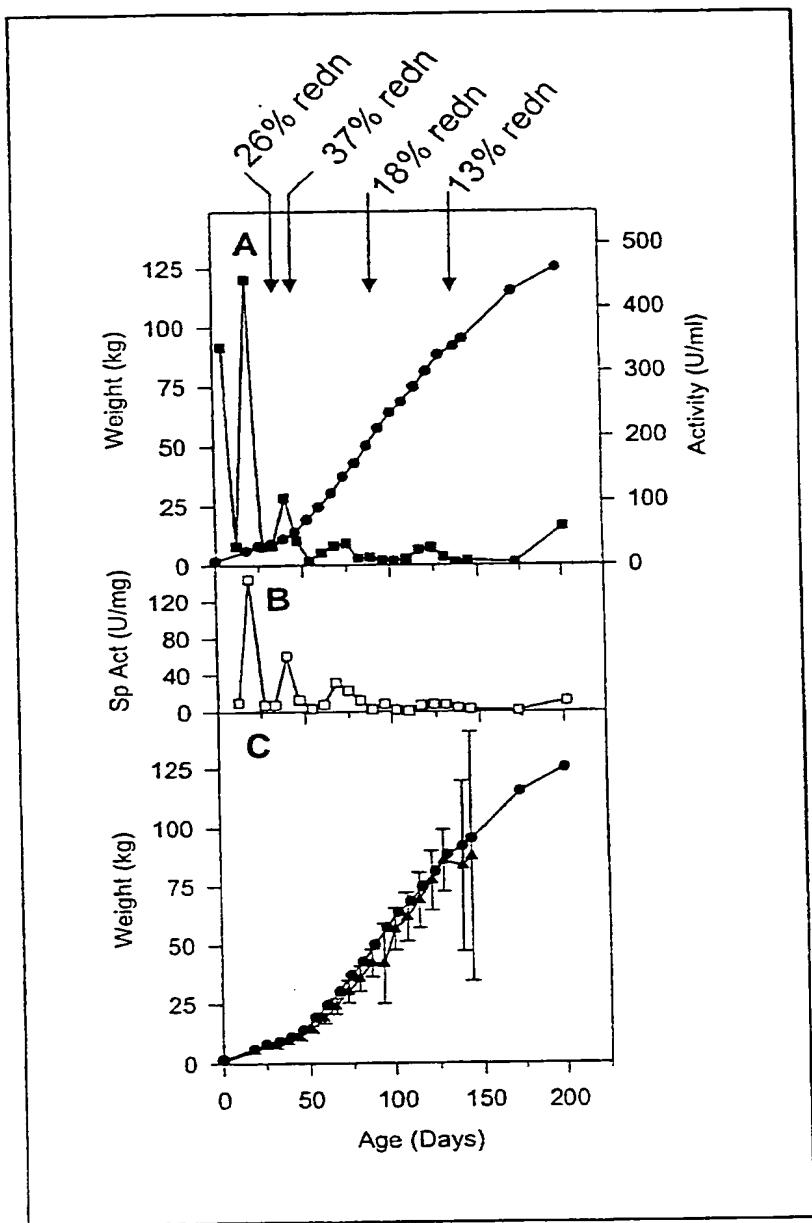


Figure 11

**Figure 12**

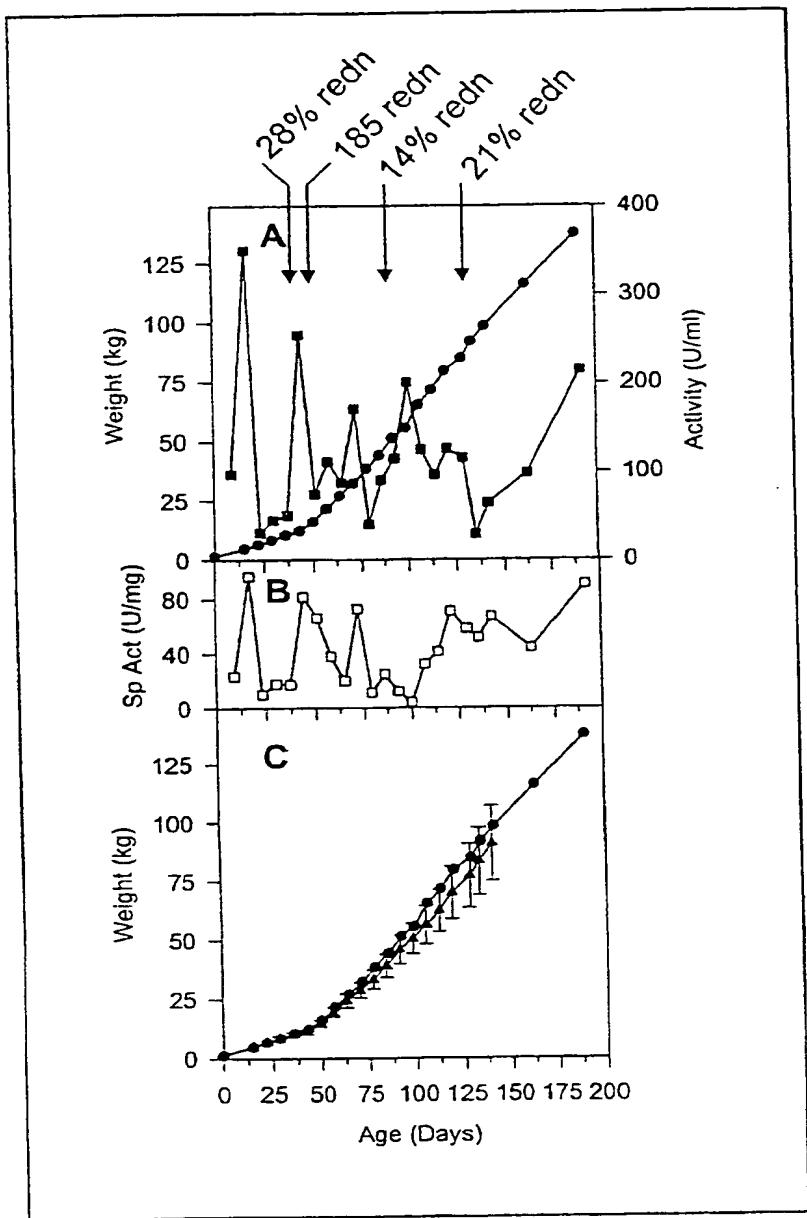


Figure 13

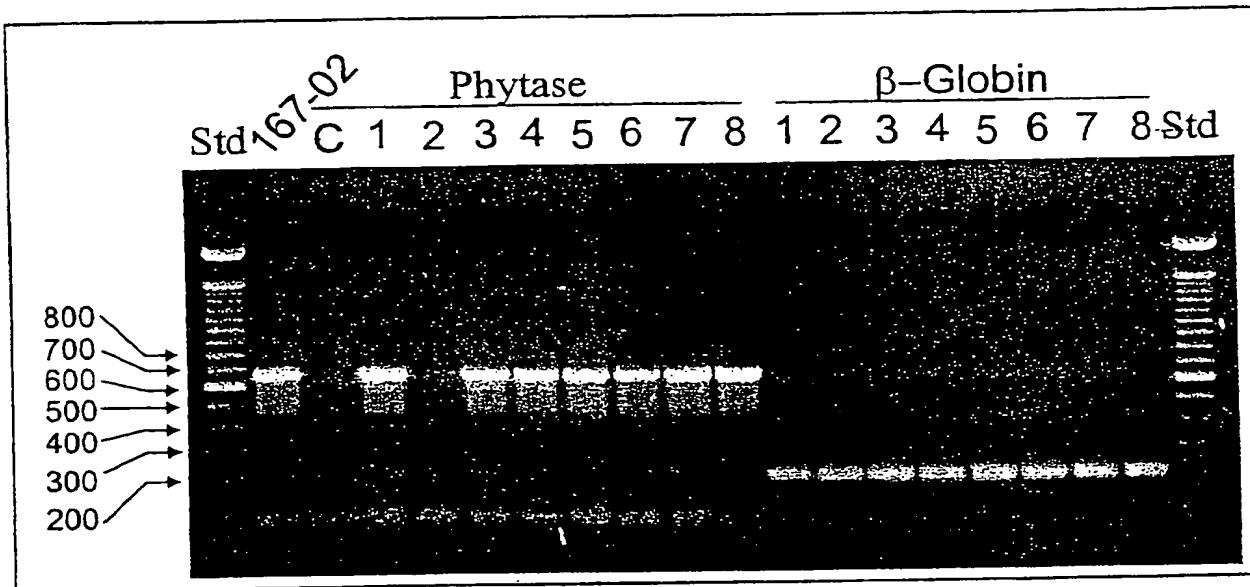


Figure 14

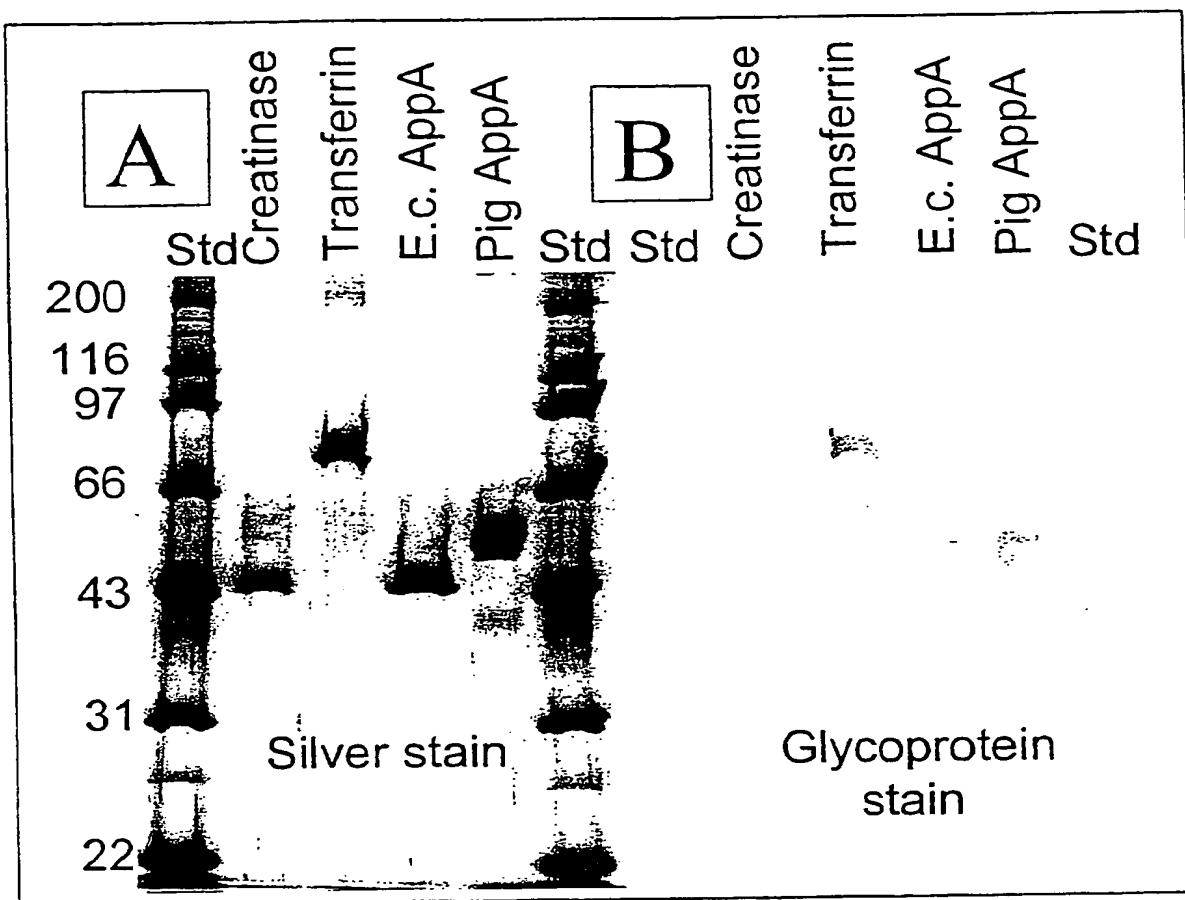


Figure 15

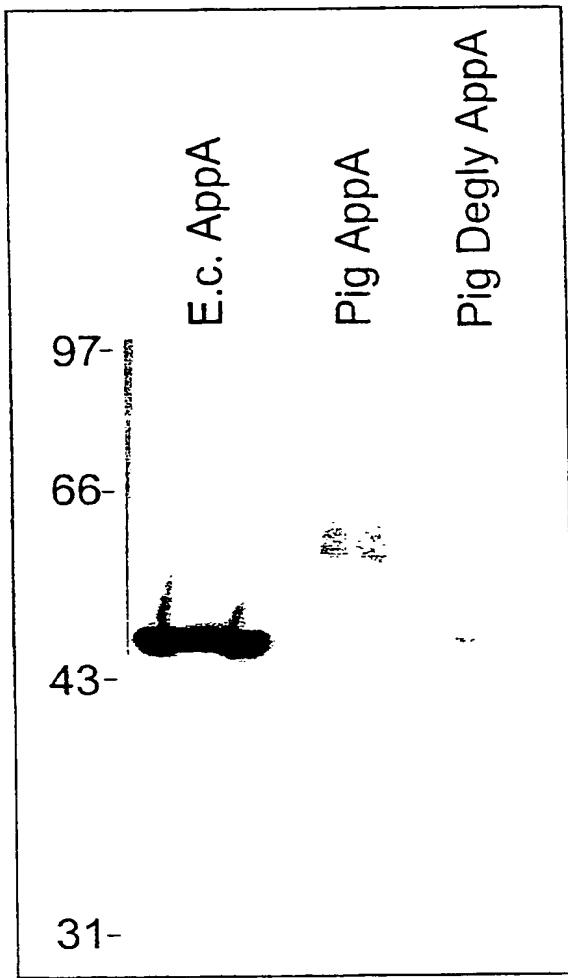


Figure 15B

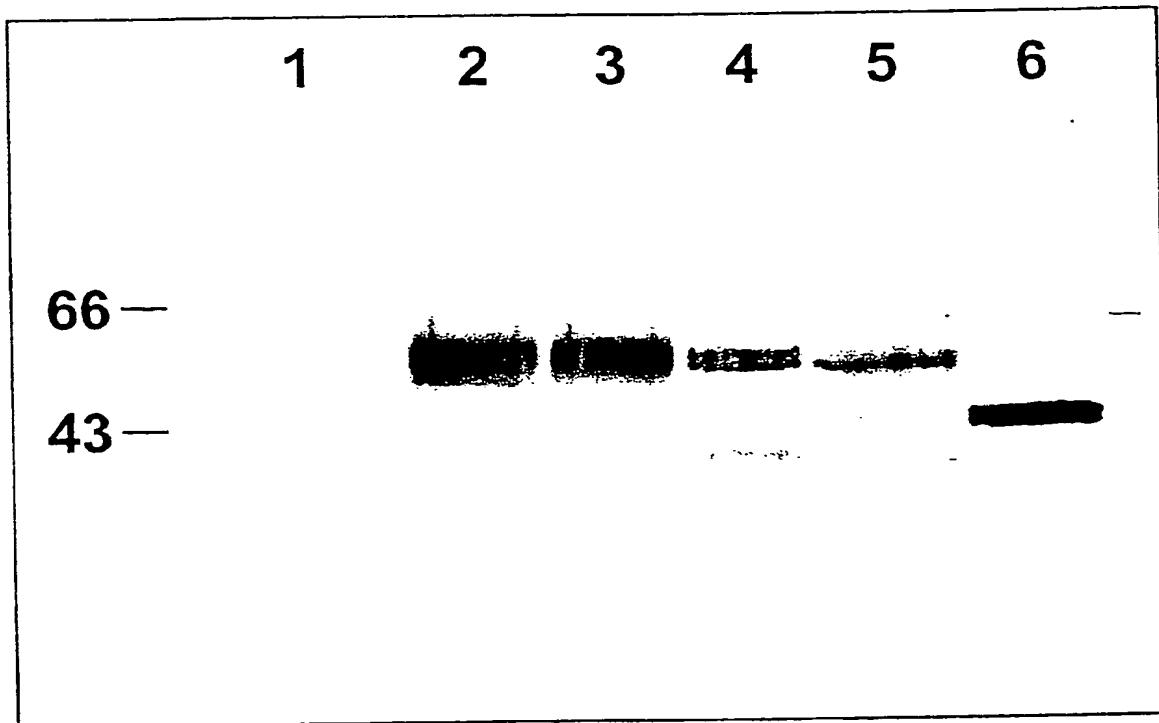


Figure 16

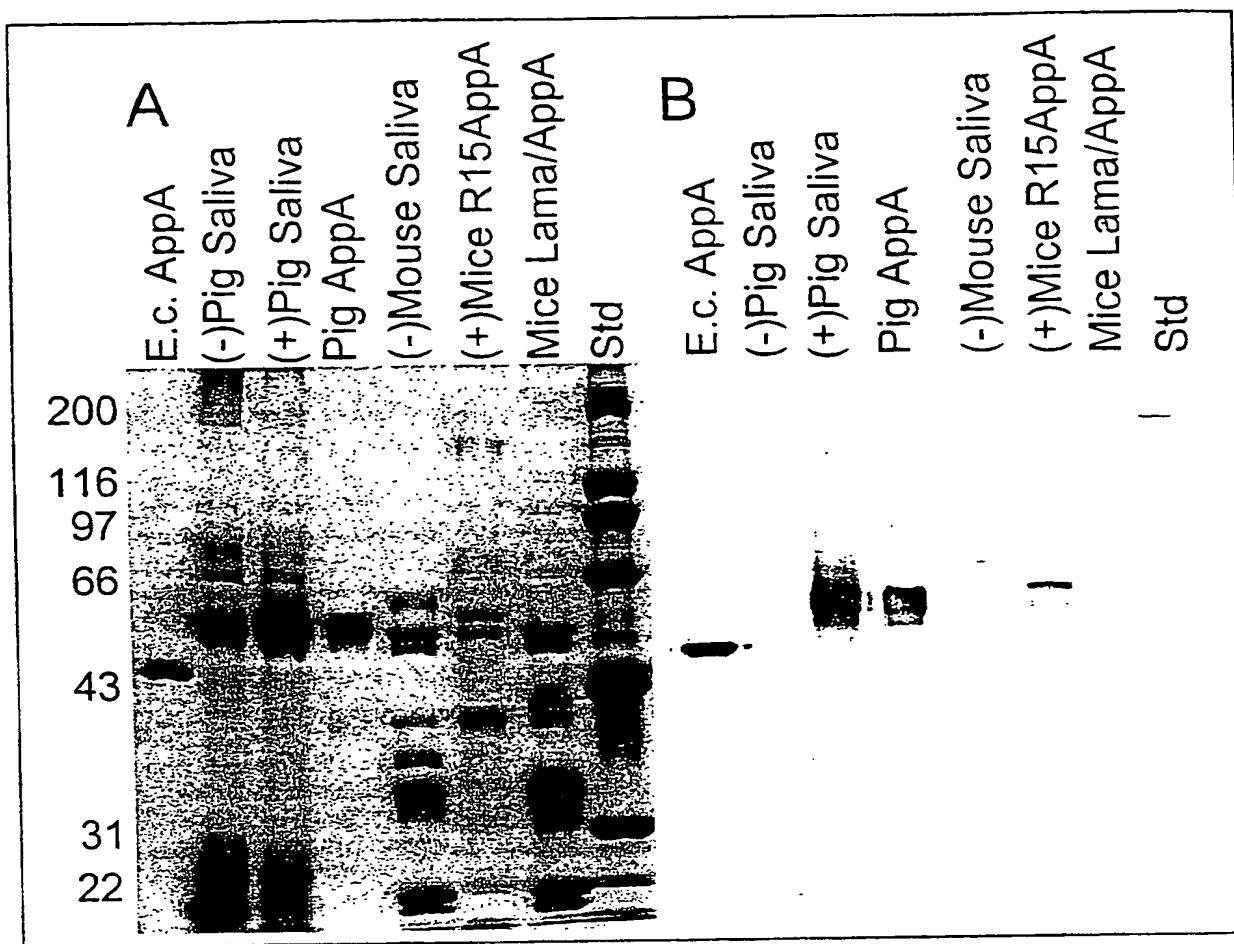


Figure 17

Figure 18: Nucleic acid sequence of the known segment of the R15/appA+intron plasmid, including the vector sequences of pBLCAT3 (SEQ ID NO:2).

LOCUS R15/appA+intron 6708 bp DNA SYN 15-APR-2000
 DEFINITION R15/appA+intron transgene with vector cut 13543 to 4954
 ACCESSION R15/appA+intron
 REFERENCE 1 (bases 1 to 6708))
 SOURCE synthetic construct.
 ORGANISM synthetic construct
 artificial sequence.
 KEYWORDS salivary proline-rich protein, acid glucose-1-phosphatase; appA
 gene; periplasmic phosphoanhydride phosphohydrolase; artificial
 sequence;
 AUTHORS Golovan, S., Forsberg, C.W., Phillips, J.
 JOURNAL Unpublished.

 DEFINITION Rat salivary proline-rich protein (RP15) gene.
 ACCESSION M64793 M36414
 VERSION M64793.1 GI:206711
 SOURCE Rat (Sprague-Dawley) liver DNA.
 ORGANISM Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata;
 Mammalia;
 Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.
 REFERENCE 1 (bases 1 to 1748)
 AUTHORS Lin, H.H. and Ann, D.K.
 TITLE Molecular characterization of rat multigene family
 encoding proline-rich proteins
 JOURNAL Genomics 10, 102-113 (1991)
 MEDLINE 91257817
 FEATURES Location/Qualifiers
 source 1..1748
 /organism="Rattus norvegicus"
 /strain="Sprague-Dawley"
 /db_xref="taxon:10116"
 /tissue_type="liver"
 /tissue_lib="cosmid genomic library"
 misc_feature 1802-1810
 /function=" consensus sequence for initiation in
 higher eukaryotes "

 FEATURES Location/Qualifiers
 DEFINITION E. coli periplasmic phosphoanhydride phosphohydrolase (appA)
 gene,
 ACCESSION M58708 L03370 L03371 L03372 L03373 L03374 L03375
 VERSION M58708.1 GI:145283
 SOURCE Escherichia coli DNA.
 ORGANISM Escherichia coli
 Bacteria; Proteobacteria; gamma subdivision;
 Enterobacteriaceae;
 Escherichia.

 REFERENCE 1 (bases 1811..3109)
 AUTHORS Dassa, J., Marck, C. and Boquet, P.L.

Figure 18 (continued):

TITLE The complete nucleotide sequence of the *Escherichia coli* gene *appA* reveals significant homology between pH 2.5 acid phosphatase and glucose-1-phosphatase
 JOURNAL *J. Bacteriol.* 172 (9), 5497-5500 (1990)
 MEDLINE 90368616

FEATURES Location/Qualifiers
 Source 1811..3109
 /organism="Escherichia coli"
 /db_xref="taxon:562"
 sig_peptide 1811.. 1876
 /gene="appA"
 CDS 1811..3109
 /gene="appA"
 /standard_name="acid phosphatase/phytase"
 /transl_table=11
 /product="periplasmic phosphoanhydride phosphohydrolase"
 /protein_id="AAA72086.1"
 /db_xref="GI:145285"

/translation="MKAILIPFLSLLIPLTPQSAFAQSEPELKLESVVIVSRHGVRAP
 TKATQLMQDVTPTDAWPTWPVKLGWLTPRGELIAYLGHYQRQRLVADGLLAKKGCPQS
 GQVAIIADVDERTRKTEAFAAGLAPDCAITVHTQADTSSPDPLFNPLKTGVCQLDNA
 NVTDAILSRAGGSIAADFTGHRQTAFRELERVLNFPQSNLCLKREKQDESCSLTQALPS
 ELKVSADNVSLTGAVSLASMLTEIFLLQQAQGMPEPGWGRITDSHQWNTLLSLHNAQF
 YLLQRTPEVARSRATPLLDLIKTAUTPHPPQKQAYGVTLPSTVLFIAQHDTNLANLGG
 ALELNWTLPQPDNTPPGELVFERWRRLSDNSQWIQVSLVFQTLQQMRDKTPLSLNT
 PPGEVVKLTLAGCEERNAQGMCSLAGFTQIVNEARIPACSL"

mat_peptide 1877 3106
 /gene="appA"
 /product="periplasmic phosphoanhydride phosphohydrolase"

mutation replace(1817.. 1819, "gcg changed to gcc")
 /gene="appA"
 /standard_name="A3 mutant"
 /note="created by site directed mutagenesis"
 /phenotype="silent mutation"

mutation replace(3092..3094, " ccg changed to ccc")
 /gene="appA"
 /standard_name=" P428 mutant"
 /note="created by site directed mutagenesis"
 /phenotype=" silent mutation "

mutation replace(3095..3097, " gcg changed to gct")
 /gene="appA"
 /standard_name=" A429 mutant"
 /note="created by site directed mutagenesis"
 /phenotype=" silent mutation "

Figure 18 (continued):

DEFINITION Plasmid pBLCAT3 (bases 3109 to 6708)

ACCESSION X64409

VERSION X64409.1 GI:58163

SOURCE synthetic construct.

ORGANISM synthetic construct

artificial sequence.

REFERENCE 1 (bases 3109 to 6708)

AUTHORS Luckow, B.H.R.

TITLE Direct Submission

JOURNAL Submitted (06-FEB-1992) B.H.R. Luckow, German Cancer Res Center, Im Neuenheimer Feld 280, W-6900 Heidelberg, FRG

REFERENCE 2 (bases 3109 to 6708)

AUTHORS Luckow, B. and Schutz, G.

TITLE CAT constructions with multiple unique restriction sites

for the functional analysis of eukaryotic promoters and regulatory elements

JOURNAL Nucleic Acids Res. 15 (13), 5490 (1987)

MEDLINE 87260024

COMMENT Promoterless CAT vector for transient transfection experiments with eukaryotic cells. Allows the analysis of foreign promoters and enhancers.

FEATURES Location/Qualifiers

source 3109 to 6116

/organism="synthetic construct"

/db_xref="taxon:32630"

SV40 t intron 3197..3810

/note="SV40 signals"

polyA_signal 3807..4047

/note="SV40 signals"

CDS complement(5244..6104)

/codon_start=1

/transl_table=11

/gene="Amp"

/product="beta-lactamase"

/protein_id="CAA45753.1"

/db_xref="GI:58165"

BASE COUNT 1916 a 1479 c 1515 g 1798 t

ORIGIN

1 GGATCCCTT TGCTATGTAG TTTTAATGG AAATTACAAC CCATAGTGTG TTGATAAATA

61 GAGAGCCTG TTTGGTTAA GCAACCTCTG TTTCTCATAA ACTCCATAAA AACAGGAATA

121 CTCTTGTGTT CTAGCATAAC CAAAGATT AGTGAATTGA AAACAATGTT CCCTTAGAGT

181 ATAGGTCTAA TAACCCGAA AATATTACCA TGATACTGAG CATTGTAAG TATCTCATAG

241 CATGTAGTAT CCATAGTCCA TCAATGAGAG AGACATTAA CATGATTTC ATTAATCAGG

301 TGGAAAAGAC ATGACAAACAT TCACAGGCAC TGCACAGAAC ATAGTGGTCC ACCTTGCACA

361 TATTCACTA AACTAGGTT ATCTATTTG TTGCTTCTC TAACATCTCT GCAATGAAGC

421 AGGTCAACAG TGCCACATAT CCTTTACTTA ACCTAAGGAA CACAAAAAAT TTTCTACATA

481 TATCCTGGTT AGAGAGTGT TAAAATAAGT TTTCCAAGAA TGGAAAAGAA ATGTTCTGAC

541 TTAACAATTA AGACAGTATT TATTTAAAGC AAGAAATATG AGGCACACAA GAAAATATTT

601 TGGGAAGAAA CCATTTGGTG AACAAATATT CAAATAAAA TAGACAAACA TAGTTAATTG

661 TAAAACATAT GTTTGACCAG CCCTTCTTT CAATAGGCTT AATGTGAATA AAATGTTAAA

721 GATTCTCTTT GGGTGGCTGC AAATTGTCCA CGAATAAGAC AAAATATAAA AATAAGGACT

781 GAGTCTCACA AAATGAAAAG GAAATATATT CAGAAAGAGA ATCTTGAGAG ATGTGTTGT

841 CACAAATTAA AGAAAACCTG TGGTGAATGA CATCCTGAGG CCTGAGCTAT TACTGACATT

Figure 18 (continued):

901 TAAGATAAAAG GTAAGCTGTAT ACATTTGTCC CATTGAGGGG ACAAGAAAGC TGCTCTCATG
 961 TTCAGCTCTA TAATTCTTGC CTTAAACAAAC TTAAATAGAA TGATTTAAAAA TATGGAGCTG
 1021 TCCATGGACC TTTGAAATAT AAAATAGTCA AGCAACTTAT CAAGGAATTAA CAGATTCCTT
 1081 GATACTAACCA CAGGTAACATC CCACACGTGT TTTGAGACTA CATTGCTGG GATTTTATTG
 1141 ATGTAATAGG TCACATGTTT TTCGGGCCAA TGTGCTGTT ATTGGTAC TTCAAGAGAA
 1201 TAGTGGCAAC TGATGCTATG TATTCTAGGG GTTTGAAGTG ATGTTTCATG ATTGAAATT
 1261 GTAAAAGAAT AACATCATCA TTCTTAACAA TAGAACATAT AAAGTCACAC AGAAGTGACA
 1321 GTGTTAACG TGTACTATTG ATCAAAGAAA TTTATTACCT TCAGTTCAA TGGAAATAAT
 1381 TACTGATAAT ACAAACATGT GTGAACACAC ACTAATCCTA TCCAAATGCA CAGTGATACA
 1441 CAGAAAATAT TAGCAAGTAG AATGCAATAT TTATATAACG ATTGTATTAA TCAATCAATT
 1501 GTATGTATCA ATATATGGGC TATTCTTA CACATGATT TATTCAAATT TACTCTAATC
 1561 ATTGTTGAAC CATTAGAAA AGGCATACTG GCAACTTTTC CTTACCTCAT CCAGCTGGC
 1621 AAAAGTCCC GTGTGGAGTA AAGGATGCAA GATTTCTGC TCTGTTAAGT ATAAAATAAT
 1681 AGTATGAATT CAAAGGTGCC ATTCTCTGC TTCTAGTTAT AAAGGCAGTG CTTGCTCTT
 1741 CCAGCACAGA TCTGGATCTC GAGGAGCTG GCGAGATT CAGGAGCTAA GGAAGCTAAA
 1801 AGCCGCCACC ATGAAAGCCA TCTTAATCCC ATTTTATCT CTTCTGATTG CGTTAACCCC
 1861 GCAATCTGCA TTCGCTCAGA GTGAGCCGGA GCTGAAGCTG GAAAGTGTGG TGATTGTCAG
 1921 TCGTCATGGT GTGCGTGCTC CAACCAAGGC CACGCAACTG ATGCAGGATG TCACCCCGA
 1981 CGCATGGCCA ACCTGGCCCG TAAAACCTGGG TTGGCTGACA CCGCGCGGTG GTGAGCTAAT
 2041 CGCCTATCTC GGACATTACC AACGCCAGCG TCTGGTAGCC GACGGATTGC TGGCGAAAAA
 2101 GGGCTGCCCG CAGTCTGGTC AGGTGCGCAT TATTGCTGAT GTCGACGAGC GTACCCGTAA
 2161 AACAGGCAGA GCCTTCGCGC CCGGGCTGGC ACCTGACTGT GCAATAACCG TACATACCA
 2221 GGCAGATACTG TCCAGTCCCC ATCCGTTATT TAATCCTCTA AAAACTGGCG TTTGCCAACT
 2281 GGATAACCGC AACGTGACTG ACACGATCCT CAGCAGGGCA GGAGGGTCAA TTGCTGACTT
 2341 TACCGGGCAT CGGCAAACCG CGTTTCGCGA ACTGGAACGG GTGCTTAATT TTCCGCAATC
 2401 AAACTTGTGC CTTAAACGTG AGAAACAGGA CGAAAGCTGT TCAATTAAACG AGGCAATTACC
 2461 ATCGGAACTC AAGGTGAGCG CCGACAATGT CTCATTAACC GGTGCGGTAA GCCTCGCATC
 2521 AATGCTGACG GAGATATTTC TCCTGCAACA AGCACAGGG AATGCCGGAGC CGGGGTGGGG
 2581 AAGGATCACC GATTACACACC AGTGGAAACAC CTTGCTAAGT TTGCTAAACG CGCAATTAA
 2641 TTTGCTACAA CGCACGCCAG AGGTTGCCCG CAGCCGCGCC ACCCCGTTAT TAGATTGAT
 2701 CAAGACAGCG TTGACGCCCG ATCCACCGCA AAAACAGGC TATGGTGTGA CATTACCCAC
 2761 TTCAGTGCTG TTTATGCCCG GACACGATAC TAATCTGGCA AATCTCGGGC CGCACTGG
 2821 GCTCAACTGG ACGCTTCCCCG GTCAGCCGGA TAACACGGCC CGAGGTGGTG AACTGGTGT
 2881 TGAACGCTGG CGTCGGCTAA GCGATAACAG CCAGTGGATT CAGGTTTCGC TGGCTTCCA
 2941 GACTTTACAG CAGATGCGTG ATAAAACGCC GCTGTCATTA AATACGCCGC CGGGAGAGGT
 3001 GAAACTGACC CTGGCAGGGAT GTGAAGAGCG AAATGCGCAG GGCATGTGTT CGTTGGCAGG
 3061 TTTTACGCAA ATCGTGAATG AAGCACGCAT ACCCGCTTGC AGTTTGTAAAG GTATAAGGCA
 3121 GTTATTGGTG CCCTTAAACG CCTGGTGCTA CGCCTGAATA AGTGATAATA AGCGGATGAA
 3181 TGGCAGAAAT TCGCCGGATC TTTGTAAGG AACCTTACTT CTGTGGTGTG ACATAATTGG
 3241 ACAAACTACC TACAGAGATT TAAAGCTCTA AGGTAATAT AAAATTAAAGTGTATAAT
 3301 GTGTTAAACT ACTGATTCTA ATTGTTGTG TATTGTTAGAT TCCAACCTAT GGAACGTGATG
 3361 AATGGGAGCA GTGGTGGAAAT GCCTTTAATG AGGAAAACCT GTTTGCTCA GAAGAAATGC
 3421 CATCTACTGA TGATGAGGCT ACTGCTGACT CTCAACATTC TACTCCTCCA AAAAAGAAGA
 3481 GAAAGGTAGA AGACCCCAAG GACTTTCTT CAGAATTGCT AAGTTTTTG AGTCATGCTG
 3541 TGTTTAGTAA TAGAACTCTT GCTGCTTTG CTATTTACAC CACAAAGGAA AAAGCTGCAC
 3601 TGCTATACAA GAAAATTATG GAAAATATT CTGTAACCTT TATAAGTAGG CATAACAGTT
 3661 ATAATCATAA CATACTGTT TTTCTTACTC CACACAGGC TAGAGTGTCT GCTATTAATA
 3721 ACTATGCTCA AAAATTGTGT ACCTTTAGCT TTTTAATTG TAAAGGGGTT AATAAGGAAT
 3781 ATTTGATGTA TAGTGCCTTG ACTAGAGATC ATAATCAGCC ATACACACATT TGTAGAGGTT
 3841 TTACTTGCTT TAAAAAAACCT CCCACACCTC CCCCTGAACC TGAAACATAA AATGAATGCA
 3901 ATTGTTGTTG TTAACCTGTT TATTGAGCT TATAATGGTT ACAAAATAAG CAATAGCATC
 3961 ACAAAATTCA CAAATAAAGC ATTTTTCTA CTGCATTCTA GTTGTGGTT GTCCAAACTC
 4021 ATCAATGTAT CTTATCATGT CTGGATCGAT CCCCGGGTAC CGAGCTCGAA TTCGTAATCA
 4081 TGGTCATAGC TGTGTTCTGT GTGAAATTGT TATCCGCTCA CAATTCCACA CAACATACGA
 4141 GCCGGAAGCA TAAAGTGAA AGCCTGGGT GCCTAATGAG TGAGCTAAT CACATTAATT
 4201 GCGTTGCGCT CACTGCCGC TTTCCAGTCG GGAAACCTGT CGTGCCAGCT GCATTAATGA
 4261 ATCGGCCAAC GCGCGGGGAG AGGCGGTTTG CGTATTGGC GCTCTCCGC TTCCTCGCTC
 4321 ACTGACTCGC TGCGCTCGGT CGTTGGCTG CGCGAGCGG TATCAGCTCA CTCAAAGGCG

Figure 18 (continued):

4381 GTAATACGGT TATCCACAGA ATCAGGGGAT AACGCAGGAA AGAACATGTG AGCAAAAGGC
 4441 CAGCAAAAGG CCAGGAACCG TAAAAAGGCC GCGTTGCTGG CGTTTTCCA TAGGCTCCGC
 4501 CCCCCCTGACG AGCATCACAA AAATCGACGC TCAAGTCAGA GGTGGCGAAA CCCGACAGGA
 4561 CTATAAAGAT ACCAGGC GTT TCCCCCTGGA AGCTCCCTCG TGCGCTCTCC TGTTCCGACC
 4621 CTGCCGCTTA CGGATACCT GTCCGCCTT CTCCCTTCGG GAAGCGTGGC GCTTCTCAA
 4681 TGCTCACGCT GTAGGTATCT CAGTCGGTG TAGTCGGTTC GCTCCAAGCT GGGCTGTGTG
 4741 CACGAACCCC CCGTTCAGCC CGACCGCTGC GCCTTATCCG GTAATATCG TCITGAGTCC
 4801 AACCCGGTAA GACACGACTT ATCGCCACTG GCAGCAGCCA CTGGTAACAG GATTAGCAGA
 4861 GCGAGGTATG TAGGCCTGTC TACAGAGTTC TTGAAGTGGT GGCCTAACTA CGGCTACACT
 4921 AGAAGGACAG TATTTGGTAT CTGCGCTCTG CTGAAGCCAG TTACCTTCGG AAAAAGAGTT
 4981 GGTAGCTCTT GATCCGGCAA ACAAAACCACC GCTGGTAGCG GTGGTTTTTG TGTTGCAAG
 5041 CAGCAGATTAA CGCGCAGAAA AAAAGGATCT CAAGAAGATC CTTTGATCTT TTCTACGGGG
 5101 TCTGACGCTC AGTGGAACGA AAACTCACGT TAAGGGATTT TGGTCATGAG ATTATCAAAA
 5161 AGGATCTTCA CCTAGATCTT TTAAATTAA AAATGAAGTT TAAATCAAT CTAAAGTATA
 5221 TATGAGTAAA CTTGGTCTGA CAGTTACCAA TGCTTAATCA GTGAGGCACC TATCTCAGCG
 5281 ATCTGTCTAT TTGTTCTATC CATAGTTGCC TGACTCCCCG TCGTGTAGAT AACTACGATA
 5341 CGGGAGGGCT TACCATCTGG CCCCAAGTGC GCAATGATAC CGCGAGACCC ACGCTCACCG
 5401 GCTCCAGATT TATCAGCAAT AAACCAAGCCA GCCGGAAGGG CCGAGCGCAG AAGTGGTCCT
 5461 GCAACTTTAT CCGCCTCCAT CCAGTCTATT AATTGTTGCC GGGAAAGCTAG AGTAAGTAGT
 5521 TCGCCAGTTA ATAGTTGCC CAACGTTGTT GCCATTGCTA CAGGCATCGT GGTGTCACGC
 5581 TCGTCGTTTG GTATGGCTTC ATTCAAGCTCC GGTTCCAAAC GATCAAGGGG AGTTACATGA
 5641 TCCCCCATGT TGTGAAAAAGCAGGTTAGC TCCCTCGGTC CTCCGATCGT TGTAGAAGT
 5701 AAGTTGGCCG CAGTGTATAC ACTCATGGTT ATGGCAGCAC TGCATAATTC TCTTACTGTC
 5761 ATGCCATCCG TAAGATGCTT TTCTGTGACT GGTGAGTACT CAACCAAGTC ATTCTGAGAA
 5821 TAGTGTATGC GGCGACCGAG TTGCTCTTGC CGGGCGTCAA TACGGGATAAA TACCGCGCCA
 5881 CATAGCAGAA CTTTAAAGT GCTCATCATT GGAAACGTT CTTGGGGCG AAAACTCTCA
 5941 AGGATCTTAC CGCTGTTGAG ATCCAGTTCG ATGTAACCCA CTCGTGCACC CAACTGATCT
 6001 TCAGCATCTT TTACTTTAC CAGCGTTCT GGGTGAGCAA AAACAGGAAG GCAAAATGCC
 6061 GCAAAAAGG GAATAAGGGC GACACGGAAA TGTTGAATAC TCATACTCTT CCTTTTCAA
 6121 TATTATTGAA GCATTATCA GGGTTATTGT CTCATGAGCG GATAACATATT TGAATGTATT
 6181 TAGAAAAATA AACAAATAGG GGTCCCGCG ACATTTCCCC GAAAAGTGCC ACCTGACGTC
 6241 TAAGAAAACCA TTATTATCAT GACATTAACC TATAAAAATA GGCGTATCAC GAGGCCCTTT
 6301 CGTCTCGCGC GTTTCGGTGA TGACGGTGAA AACCTCTGAC ACATGCAGCT CCCGGAGACG
 6361 GTCACAGCTT GTCTGTAAGC GGATGCCGGG AGCAGACAAG CCCGTCAGGG CGCGTCAGCG
 6421 GGTGTTGGCG GGTGTCGGGG CTGGCTTAAC TATGCGGCAT CAGACGAGAT TGTACTGAGA
 6481 GTGCACCATA TGCGGTGTGA AATACCGCAC AGATGCGTAA GGAGAAAATA CCGCATCAGG
 6541 CGCCATTGCG CATTCAAGCT GCGCAACTGT TGGGAAGGGC GATCGGTGCG GGCCTCTTCG
 6601 CTATTACGCC AGCTGGCGAA AGGGGGATGT GCTGCAAGGC GATTAAGTTG GTAAACGCCA
 6661 GGGTTTCCC AGTCACGACG TTGTAAAACG ACGGCCAGTG CCAAGCTT

//

Figure 19: Nucleic acid sequence of the known segment of the R15/appa+intron transgene used for the generation of transgenic mice (SEQ ID NO: 3).

LOCUS R15/appa 4060 bp DNA SYN 15-APR-2000
 DEFINITION R15/appa transgene without vector
 ACCESSION R15/appa
 REFERENCE 1 (bases 1 to 4060)
 SOURCE synthetic construct.
 ORGANISM synthetic construct
 artificial sequence.
 KEYWORDS salivary proline-rich protein, acid glucose-1-phosphatase; appA gene; periplasmic phosphoanhydride phosphohydrolase; artificial sequence;
 AUTHORS Golovan, S., Forsberg, C.W., Phillips, J.
 JOURNAL Unpublished.

DEFINITION Rat salivary proline-rich protein (RP15) gene.
 ACCESSION M64793 M36414
 VERSION M64793.1 GI:206711
 SOURCE Rat (Sprague-Dawley) liver DNA.
 ORGANISM Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata;
 Mammalia;
 Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.
 REFERENCE 1 (bases 1 to 1748)
 AUTHORS Lin, H.H. and Ann, D.K.
 TITLE Molecular characterization of rat multigene family
 encoding proline-rich proteins
 JOURNAL Genomics 10, 102-113 (1991)
 MEDLINE 91257817
 FEATURES Location/Qualifiers
 source 1..1748
 /organism="Rattus norvegicus"
 /strain="Sprague-Dawley"
 /db_xref="taxon:10116"
 /tissue_type="liver"
 /tissue_lib="cosmid genomic library"
 misc_feature 1802-1810
 /function=" consensus sequence for initiation in
 higher eukaryotes "

FEATURES Location/Qualifiers
 DEFINITION E. coli periplasmic phosphoanhydride phosphohydrolase (appA) gene,
 ACCESSION M58708 L03370 L03371 L03372 L03373 L03374 L03375
 VERSION M58708.1 GI:145283
 SOURCE Escherichia coli DNA.
 ORGANISM Escherichia coli
 Bacteria; Proteobacteria; gamma subdivision;
 Enterobacteriaceae;
 Escherichia.
 REFERENCE 1 (bases 1811..3109)
 AUTHORS Dassa, J., Marck, C. and Boquet, P.L.

Figure 19 (continued):

TITLE The complete nucleotide sequence of the *Escherichia coli* gene *appA* reveals significant homology between pH 2.5 acid phosphatase and glucose-1-phosphatase
 JOURNAL *J. Bacteriol.* 172 (9), 5497-5500 (1990)
 MEDLINE 90368616

FEATURES **Location/Qualifiers**
 Source 1811..3109
 /organism="Escherichia coli"
 /db_xref="taxon:562"
 sig_peptide 1811..1876
 /gene="appA"
 CDS 1811..3109
 /gene="appA"
 /standard_name="acid phosphatase/phytase"
 /transl_table=11
 /product="periplasmic phosphoanhydride phosphohydrolase"
 /protein_id="AAA72086.1"
 /db_xref="GI:145285"

 /translation="MKAILIPFLSLLIPLTPQSAFAQSEPELKLESVVIVSRHGVRAP
 TKATQLMQDVTPTDAWPTWPVKLGWLTPRGGELIAYLGHYQRQRLVADGLLAKKGCPQS
 GQVAIIDVDERTRKTGEAFAAGLAPDCAITVHTQADTSSPDPLFNPLKTGVCQLDNA
 NVTDAILSRAGGSIADFTGHRQTAFRELERVLNFQSNLCLKREKQDESCSLTQALPS
 ELKVSADNVSLTGAWSLASMLTEIFLLQQAQGMPEPGWGRITDSHQWNTLLSLHNAQF
 YLLQRTPEVARSRATPLLDLIKALTYPHPQKQAYGVTLPSTSVLFIAGHDTNLANLGG
 ALELNWTLPQPDNTPPGELVFERWRRLSDNSQWIQVSLVFQTLQQMRDKTPLSLNT
 PPGEVKTLAGCEERNAQGMCSLAGFTQIVNEARIPACSL"
 mat_peptide 1877 3106
 /gene="appA"
 /product="periplasmic phosphoanhydride phosphohydrolase"

 mutation replace(1817.. 1819, "gca changed to tca")
 /gene="appA"
 /standard_name="A3 mutant"
 /note="created by site directed mutagenesis"
 /phenotype="silent mutation"
 mutation replace(3092..3094, "ccg changed to ccc")
 /gene="appA"
 /standard_name=" P428 mutant"
 /note="created by site directed mutagenesis"
 /phenotype=" silent mutation "
 mutation replace(3095..3097, "gca changed to tca")
 /gene="appA"
 /standard_name=" A429 mutant"
 /note="created by site directed mutagenesis"
 /phenotype=" silent mutation "

Figure 19 (continued):

```

SV40 t intron 3197..3810
                /note="SV40 signals"
polyA_signal    3807..4047
                /note="SV40 signals"

```

BASE COUNT 1257 a 814 c 843 g 1146 t
 ORIGIN

```

1 GGATCCCCCT TGCTATGTAG TTTTTAATGG AAATTACAAC CCATAGTGTG TTGATAAATA
61 GAGAGTCCTG TTTGGTTAA GCAACCTCTG TTTCTCATAA ACTCCATAAA AACAGGAATA
121 CTCTTGTCTT CTAGCATAAC CAAAAGATT AGTGAATTGA AAACAATGTT CCCTTAGAGT
181 ATAGGTCTAA TAACCCCGAA AATATTACCA TGATACTGAG CATTGTAAAG TATCTCATAG
241 CATGTAGTAT CCATAGTCCA TCAATGAGAG AGACATTAA CATGATTTTC ATTAATCAGG
301 TGGAAAAGAC ATGACAAACAT TCACAGGCAC TGACAGAAC ATAGTGGTCC ACCTTGCACA
361 TATTTCACTA AACTAGGTTT ATCTATTTG TTGCTTTCT TAACATCTCT GCAATGAAGC
421 AGGTCAACAG TGCCACATAT CCTTTACTTA ACCTAAGGAA CACAAAAAAAT TTTCTACATA
481 TATCCTGGTT AGAGAGTGT TAAAATAAGT TTCCAAGAA TGGAAAAGAA ATGTTCTGAC
541 TTAACAATTA AGACAGTATT TATTTAAAGC AAGAAATATG AGGCACACAA GAAAATATTT
601 TGGGAAGAAA CCATTTGGTG AACAAATATT CAAATAAAAA TAGACAAACAA TAGTTAATTG
661 TAAAACATAT GTTGACCAG CCCTTCTTTT CAATAGGTT AATGTGAATA AAATGTTAAA
721 GATTCTCTT GGGTGGCTGC AAATTGTCCA CAAAGAGAGA ATCTTGAGAG AATGTTGTGT
781 GAGTCTCACA AAATGAAAAG GAAATATATT CATCCTGAGG CCTGAGCTAT TACTGACATT
841 CACAAATTAA AGAAAACCTG TGGTGAATGA CATTGAGGGG ACAAGAAAGC TGCTCTCATG
901 TAAGATAAAAG GTAACTGTAT ACATTGTCC TTAAATAGAA TGATTTAAAAA TATGGAGCTG
961 TTCAGCTCTA TAATTCTTGC CTTAAACAAAC AGCAACTTAT CAAGGAATTA CAGATCCTT
1021 TCCATGGACC TTTGAAATAT AAAATAGTCA 1081 GATACTAACAA CAGGTAAATC CCACACGTGT
1141 ATGTAATAGG TCACATGTTT TTCGGGCCAA 1201 TAGTGGCAAC TGATGCTATG TATTCTAGGG
1261 GTAAAAGAAT AACATCATCA TTCTTAACAA 1321 GTGTTAACG TGTACTATTG ATCAAAGAAA
1381 TACTGATAAT ACAAACATGT GTAACACAC 1441 CAGAAAATAT TAGCAAGTAG AATGCAATAT
1501 GTATGTATCA ATATATGGGC TATTTTCTTA 1561 ATTGTTGAAC CATTAGAAA AGGCATACTG
1621 AAAAGTCCCAGTGTGGAGTA AAGGATGCAA 1681 AGTATGAATT CAAAGGTGCC ATTCTTCTGC
1741 CCAGCACAGA TCTGGATCTC GAGGAGCTTG 1801 AGCCGCCACC ATGAAAGCCA TCTTAATCCC
1861 GCAATCTGCA TTGCGCTCAGA GTGAGCCGG 1921 TCGTCATGGT GTGCGTGCCTC CAACCAAGGC
1981 CGCATGGCCA ACCTGGCCGG TAAAAGTGGG 2041 CGCCTATCTC GGACATTACC AACGCCAGCG
2101 GGGCTGCCCG CAGTCTGGTC AGGTGCGAT 2161 AACAGGGCAGA GCCTTCGCCGG CGGGCTGGC
2221 GGCAGATACG TCCAGTCCCG ATCCGTTATT 2281 GGATAACGCG AACGTGACTG ACGCGATCCT
2341 TACCGGGCAT CGGCAAACGG CGTTTGCAGA 2401 AAAACTGTGC CTTAACACGT AGAAACAGGA
2461 ATCGGAAACTC AAGGTGAGCG CCGACAATGT 2521 AATGCTGACG GAGATATTTC TCCTGCAACA
2581 AAGGATCACC GATTACACCC AGTGGAACAC 2641 TTTGCTACAA CGCACGCCAG AGGTTGCCGG
2701 CAAGACAGCG TTGACGCCCG ATCCACCGCA 2761 TTCAGTGTG 2821 GCTCAACTGG ACGCTTCCCG GTCAAGCCGG
2881 TGAACGCTGG CGTCGGCTAA GCGATAACAG 2881 TGAACGCTGG CGTCGGCTAA GCGATAACAG

```

Figure 19 (continued):

2941 GACTTACAG CAGATGCGTG ATAAAACGCC GCTGTCATTA AATACGCCGC CCGGAGAGGT
3001 GAAACTGACC CTGGCAGGAT GTGAAGAGCG AAATGCCAG GGCATGTGTT CGTTGGCAGG
3061 TTTTACGCAA ATCGTGAATG AAGCACGCAT ACCCGCTTGC AGTTTGTAAAG GTATAAGGCA
3121 GTTATTGGTG CCCTTAAACG CCTGGTGCTA CGCCTGAATA AGTGATAATA AGCGGATGAA
3181 TGGCAGAAAT TCGCCGGATC TTTGTGAAGG AACCTTACTT CTGTGGTGTG ACATAATTGG
3241 ACAAACTACC TACAGAGATT TAAAGCTCTA AGGTAAATAT AAAATTTTTA AGTGTATAAT
3301 GTGTTAAACT ACTGATTCTA ATTGTTGTG TATTTTAGAT TCCAACCTAT GGAACGTGATG
3361 AATGGGAGCA GTGGTGAAT GCCTTAAATG AGGAAAACCT GTTTGCTCA GAAGAAATGC
3421 CATCTAGTGA TGATGAGGCT ACTGCTGACT CTCAACATTC TACTCCTCCA AAAAAGAAGA
3481 GAAAGGTAGA AGACCCCAAG GACTTTCCCTT CAGAATTGCT AAGTTTTTG AGTCATGCTG
3541 TGTTTAGTAA TAGAACTCTT GCTTGCTTG CTATTTACAC CACAAAGGAA AAAGCTGCAC
3601 TGCTATACAA GAAAATTATG GAAAATATT CTGTAACCTT TATAAGTAGG CATAACAGTT
3661 ATAATCATAA CATACTGTCTT TTTCTTACTC CACACAGGCA TAGAGTGTCT GCTATTAATA
3721 ACTATGCTCA AAAATTGTGT ACCTTTAGCT TTTTAATTG TAAAGGGGTT AATAAGGAAT
3781 ATTTGATGTA TAGTGCCTTG ACTAGAGATC ATAATCAGCC ATACCACATT TGTAGAGGTT
3841 TTACTTGCTT TAAAAAACCT CCCACACCTC CCCCTGAACC TGAACATAA AATGAATGCA
3901 ATTGTTGTTG TTAACCTGTT TATTGCAGCT TATAATGGTT ACAAAATAAG CAATAGCATC
3961 ACAAAATTCA CAAATAAAAGC ATTTTTTCA CTGCATTCTA GTTGTGGTTT GTCCAAACTC
4021 ATCAATGTAT CTTATCATGT CTGGATCGAT CCCCCGGGTAC

//

Figure 20: Nucleic acid sequence of the known segment of the R15/appa plasmid (including the vector sequences of pBLCAT3 (SEQ ID NO:4).

LOCUS R15/appa 6116 bp DNA **SYN** 15-APR-2000
DEFINITION R15/appa transgene with vector
ACCESSION R15/appa
REFERENCE 1 (bases 1 to 6116)
SOURCE synthetic construct.
ORGANISM synthetic construct
artificial sequence.
KEYWORDS salivary proline-rich protein, acid glucose-1-phosphatase; appA
gene; periplasmic phosphoanhydride phosphohydrolase; artificial
sequence;
AUTHORS Golovan, S., Forsberg, C.W., Phillips, J.
JOURNAL Unpublished.

DEFINITION Rat salivary proline-rich protein (RP15) gene.
ACCESSION M64793 M36414
VERSION M64793.1 GI:206711
SOURCE Rat (Sprague-Dawley) liver DNA.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata;
Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 1748)
AUTHORS Lin, H.H. and Ann, D.K.
TITLE Molecular characterization of rat multigene family
encoding proline-rich proteins
JOURNAL Genomics 10, 102-113 (1991)
MEDLINE 91257817
FEATURES Location/Qualifiers
source 1..1748
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/tissue_type="liver"
/tissue_lib="cosmid genomic library"
misc_feature 1802-1810
/function=" consensus sequence for initiation in
higher eukaryotes "

FEATURES Location/Qualifiers
DEFINITION E. coli periplasmic phosphoanhydride phosphohydrolase (appA)
gene,
ACCESSION M58708 L03370 L03371 L03372 L03373 L03374 L03375
VERSION M58708.1 GI:145283
SOURCE Escherichia coli DNA.
ORGANISM Escherichia coli
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.

REFERENCE 1 (bases 1811..3109)
AUTHORS Dassa, J., Marck, C. and Boquet, P.L.
TITLE The complete nucleotide sequence of the Escherichia coli gene appA
reveals significant homology between pH 2.5 acid phosphatase
and glucose-1-phosphatase
JOURNAL J. Bacteriol. 172 (9), 5497-5500 (1990)

Figure 20 (continued):

MEDLINE 90368616

FEATURES Location/Qualifiers

Source 1811..3109
 /organism="Escherichia coli"
 /db_xref="taxon:562"

sig_peptide 1811..1876

CDS 1811..3109
 /gene="appA"
 /standard_name="acid phosphatase/phytase"
 /transl_table=11
 /product="periplasmic phosphoanhydride phosphohydrolase"
 /protein_id="AAA72086.1"
 /db_xref="GI:145285"

/translation="MKAILIPFLSLIPLTPQSAFAQSEPELKLESVVIVSRHGVRAP
 TKATQLMODVTPDAWTPVVKLGWLTPRGELIAYLGHYQRQRLVADGLLAKKGCPQS
 GQVAIIDVDERTRKTGEAFAAGLAPDCAITVHTQADTSSPDPLFNPLKTGVCQLDNA
 NVTDAILSRAGGSIADFTGHRQTAFRELERVLNPQSNLCLKREKQDESCSLTQALPS
 ELKVSADNVSLTGAVSLASMLTEIFLLQQAQGMPEPGWGRITDSHQWNTLLSLHNAQF
 YLLQRTPEVARSRATPLLDLIKTAITPHPPQKQAYGVTLPSTSVLFIAGHDTNLANLGG
 ALELNWTLPQPDNTPPGELVFERWRLSDNSQWIQVSLVQFTLQQMRDKTPLSLNT
 PPGEVKLTLAGCEERNAQGMCSLAGFTQIVNEARIIPACSL"

mat_peptide 1877 3106
 /gene="appA"
 /product="periplasmic phosphoanhydride phosphohydrolase"

mutation replace(1817..1819, "gcg changed to gcc")
 /gene="appA"
 /standard_name="A3 mutant"
 /note="created by site directed mutagenesis"
 /phenotype="silent mutation"

mutation replace(3092..3094, "ccg changed to ccc")
 /gene="appA"
 /standard_name=" P428 mutant"
 /note="created by site directed mutagenesis"
 /phenotype=" silent mutation "

mutation replace(3095..3097, "gcg changed to gct")
 /gene="appA"
 /standard_name=" A429 mutant"
 /note="created by site directed mutagenesis"
 /phenotype=" silent mutation "

DEFINITION Plasmid pBLCAT3 (bases 3109 to 6116)

ACCESSION X64409

VERSION X64409.1 GI:58163

SOURCE synthetic construct.

ORGANISM synthetic construct

artificial sequence.

REFERENCE 1 (bases 3109 to 6116)

AUTHORS Luckow, B.H.R.

TITLE Direct Submission

JOURNAL Submitted (06-FEB-1992) B.H.R. Luckow, German Cancer Res Center, Im Neuenheimer Feld 280, D-6900 Heidelberg, FRG

Figure 20 (continued):

REFERENCE 2 (bases 3109 to 6116)
 AUTHORS Luckow, B. and Schutz, G.
 TITLE CAT constructions with multiple unique restriction sites
 for the functional analysis of eukaryotic promoters and
 regulatory elements
 JOURNAL Nucleic Acids Res. 15 (13), 5490 (1987)
 MEDLINE 87260024
 COMMENT Promoterless CAT vector for transient transfection
 experiments with eukaryotic cells. Allows the analysis of foreign
 promoters and enhancers.
 FEATURES Location/Qualifiers
 source 3109 to 6116
 /organism="synthetic construct"
 /db_xref="taxon:32630"
 polyA_signal 3262..3457
 /note="SV40 signals"
 CDS complement(4654..5514)
 /codon_start=1
 /transl_table=11
 /gene="Amp"
 /product="beta-lactamase"
 /protein_id="CAA45753.1"
 /db_xref="GI:58165"
 BASE COUNT 1724 a 1386 c 1407 g 1599 t
 ORIGIN
 1 GGATCCCCTT TGCTATGTAG TTTTAATGG AAATTACAAC CCATAGTGTG TTGATAAATA
 61 GAGAGTCCTG TTTGGTTAA GCAACCTCTG TTTCTCATAA ACTCCATAAA AACAGGAATA
 121 CTCTTGTCTT CTAGCATAAC CAAAAGATT AGTGAATTGA AAACAATGTT CCCTTAGAGT
 181 ATAGGTCTAA TAACCCCGAA AATATTACCA TGATACTGAG CATTGTAAG TATCTCATAG
 241 CATGTAGTAT CCATAGTCCA TCAATGAGAG AGACATTTAA CATGATTTTC ATTAATCAGG
 301 TGGAAAAGAC ATGACAACAT TCACAGGCAC TGCACAGAAC ATAGTGGTCC ACCTTGCACA
 361 TATTCACTA AACTAGGTTT ATCTATTTG TTGCTTCTC TAACATCTCT GCAATGAAGC
 421 AGGTCAACAG TGCCACATAT CCTTTACTTA ACCTAAGGAA CACAAAAAAT TTTCTACATA
 481 TATCCTGGTT AGAGAGTGCT TAAAATAAGT TTTCCAAGAA TGGAAAAGAA ATGTTCTGAC
 541 TTAACAATTAA AGACAGTATT TATTTAAAGC AAGAAATATG AGGCACACAA GAAAATATTT
 601 TGGGAAGAAA CCATTTGGTG AACAATATTT CAAATAAAA TAGACAAACA TAGTTAATTG
 661 TAAAACATAT GTTGACCAG CCCTTCTTTT CAATAGGCTT AATGTGAATA AAATGTTAAA
 721 GATTCTCTT GGGTGGCTGC AAATTGTCCA CGAATAAGAC AAAATATAAA AATAAGGACT
 781 GAGTCTCACA AAATGAAAAG GAAATATATT CAGAAAGAGA ATCTTGAGAG AATGTTGTG
 841 CACAAATTAA AGAAAACCTG TGGTGAATGA CATCCTGAGG CCTGAGCTAT TACTGACATT
 901 TAAGATAAAAG GTAACTGTAT ACATTTGTCC CATTGAGGG ACAAGAAAGC TGCTCTCATG
 961 TTCAGCTCTA TAATTCTTGC CTTAAACAAAC TTAAATAGAA TGATTTAAA TATGGAGCTG
 1021 TCCATGGACC TTTGAAATAT AAAATAGTCA AGCAACTTAT CAAGGAATTAA CAGATTCTT
 1081 GATACTAACAA CAGGTAAATC CCACACGTGT TTTGAGACTA CATTGCTGG GATTTTATTG
 1141 ATGTAATAGG TCACATGTTT TCGGGCCAA TGTTGCTGTT ATTGCGTTAC TTCAAGAGAA
 1201 TAGTGGCAAC TGATGCTATG TATTCTAGGG GTTGAAAGTG ATGTTTCATG ATTGAAATT
 1261 GTAAAAGAAT AACATCATCA TTCTTAACAA TAGAACATAT AAAGTCACAC AGAAGTGACA
 1321 GTGTTAACG TGTACTATTG ATCAAAGAAA TTTATTACCT TCAGTTCAA TGGAATAAT
 1381 TACTGATAAT ACAAAACATGT GTGAACACAC ACTAACCTCA TCCAAATGCA CAGTGATAACA
 1441 CAGAAAATAT TAGCAAGTAG AATGCAATAT TTATATAACG ATTGTATTAA TCAATCAATT
 1501 GTATGTATCA ATATATGGC TATTTCTTA CACATGATT TATTCAAATT TACTCTAATC
 1561 ATTGTTGAAC CATTAGAAA AGGCATACTG GCAACTTTTC CTTACCTCAT CCAGCTGGC
 1621 AAGAGTCCCA GTGTGGAGTA AAGGATGCAA GATTTCCTGC TCTGTTAAGT ATAAAATAAT

Figure 20 (continued):

1681 AGTATGAATT CAAAGGTGCC ATTCTTCTGC TTCTAGTTAT AAAGGCAGTG CTTGCTTCTT
 1741 CCAGCACAGA TCTGGATCTC GAGGAGCTG GCGAGATTT CAGGAGCTAA GGAAGCTAA
 1801 AGCCGCCACC ATGAAAGCCA TCTTAATCCC ATTTTATCT CTTCTGATTG CGTTAACCCC
 1861 GCAATCTGCA TTCGCTCAGA GTGAGCCGGA GCTGAAGCTG GAAAGTGTGG TGATTGTCAG
 1921 TCGTCATGGT GTGCGTGCTC CAACCAAGGC CACGCAACTG ATGCAGGATG TCACCCCAGA
 1981 CGCATGGCCA ACCTGGCCGG TAAAACCTGGG TTGGCTGACA CCGCGCGGTG GTGAGCTAAT
 2041 CGCCTATCTC GGACATTAC AACGCCAGCG TCTGGTAGCC GACGGATTGC TGGCGAAAAAA
 2101 GGGCTGCCCG CAGTCTGGTC AGGTCGCGAT TATTGCTGAT GTCGACGAGC GTACCCGTAA
 2161 AACAGGGGAA GCCTTCGCGG CCGGGCTGGC ACCTGACTGT GCAATAACCG TACATACCCA
 2221 GGCAGATACG TCCAGTCCCCG ATCCGTTATT TAATCCTCTA AAAACTGGCG TTTGCCAAC
 2281 GGATAACGCG AACGTGACTG ACGCGATCCT CAGCAGGGCA GGAGGGTCAA TTGCTGACTT
 2341 TACCGGGCAT CGGCAACAGG CGTTTCGCGA ACTGGAACGG GTGCTTAATT TTCCGCAATC
 2401 AAACTTGTGC CTTAAACGTG AGAAACAGGA CGAAAGCTGT TCATTAACGC AGGCATTACC
 2461 ATCGGAACTC AAGGTGAGCG CCGACAATGT CTCATTAACC GGTGCGGTAA GCCTCGCATC
 2521 AATGCTGACG GAGATATTTC TCCTGCAACA AGCACAGGGAA ATGCGGGAGC CGGGGTGGGG
 2581 AAGGATCACC GATTCACACC AGTGGAACAC CTTGCTAAGT TTGCTAAACG CGCAATTAA
 2641 TTTGCTACAA CGCACGCCAG AGGTTGCGCG CAGCCGCGCC ACCCCGTTAT TAGATTGAT
 2701 CAAGACAGCG TTGACGCCCG ATCCACCGCA AAAACAGGGC TATGGTGTGA CATTACCCAC
 2761 TTCAGTGCTG TTTATCGCCG GACACGATAC TAATCTGGCA AATCTCGGCG GCGCACTGGA
 2821 GCTCAACTGG ACGCTTCCCG GTCAAGCCGG A TAACACGCCG CCAGGTGGTG AACTGGTGT
 2881 TGAACGCTGG CGTCGGCTAA GCGATAACAG CCAGTGGATT CAGGTTTCGC TGGTCTTCCA
 2941 GACTTACAG CAGATGCGTG ATAAAACGCC GCTGTCATTA AATACGCCGC CCGGAGAGGT
 3001 GAAACTGACC CTGGCAGGAT GTGAAGAGCG AAATGCGCAG GGCATGTGTT CGTTGGCAGG
 3061 TTTTACGCAA ATCGTGAATG AAGCACGCAT ACCCGCTTGC AGTTTGTAAAG GTATAAGGC
 3121 GTTATTGGTG CCCTTAAACG CCTGGTGCTA CGCCTGAATA AGTGATAATA AGCGGATGAA
 3181 TGGCAGAAAT TCGCCGGATC TTTGTGAAGG AACCTTAACCTT CTGTGGTGTG ACATAATTGG
 3241 ACAAAACTACC TACAGAGATT TAAAAAAACCT CCCACACCTC CCCCTGAACC TGAAACATAA
 3301 AATGAATGCA ATTGTTGTTG TTAACCTGTT TATTGCACT TATAATGGTT ACAAATAAAG
 3361 CAATAGCATE ACAAAATTCA CAAATAAAGC ATTTTTTCA CTGCATTCTA GTTGTGGTT
 3421 GTCCAAACTC ATCAATGTAT CTTATCATGT CTGGATCGAT CCCCGGGTAC CGAGCTCGAA
 3481 TTCGTAATCA TGGTCATAGC TGTTCCTGT GTGAAATTGT TATCCGCTCA CAATTCCACA
 3541 CAACATACGA GCCGGAAAGCA TAAAGTGTAA AGCCTGGGT GCCTAATGAG TGAGCTAACT
 3601 CACATTAATT GCGTTGCGCT CACTGCCGC TTTCCAGTCG GGAAACCTGT CGTGCCAGCT
 3661 GCATTAATGA ATCGGCCAAC GCGCGGGGAG AGGCGTTTG CGTATTGGGC GCTCTTCGCG
 3721 TTCCTCGCTC ACTGACTCGC TGCCTCGGT CGTTCGCTG CGCGGAGCGG TATCAGCTCA
 3781 CTCAAAGGCG GTAATACGGT TATCCACAGA ATCAGGGGAT AACGCAGGAA AGAACATGTG
 3841 AGCAAAAGGC CAGCAAAAGG CCAGGAACCG TAAAAGGCC GCGTTGCTGG CGTTTTTCCA
 3901 TAGGCTCCGC CCCCCTGACG AGCATCACAA AAATGACGC TCAAGTCAGA GGTGGCGAAA
 3961 CCCGACAGGA CTATAAAGAT ACCAGGCCTT TCCCCCTGGA AGCTCCCTCG TGCGCTCTCC
 4021 TGTTCCGACC CTGCCGCTTA CCGGATACCT GTCCGCTTT CTCCCTTCCG GAAGCGTGGC
 4081 GCTTCTCAA TGCTCACGCT GTAGGTATCT CAGTTCGGT TAGGTCGTT GCTCCAAGCT
 4141 GGGCTGTGTG CACGAACCCC CGTTCAGCC CGACCGCTGC GCCTTATCCG GTAACATATCG
 4201 TCTTGAGTCC AACCCGGTAA GACACGACTT ATGCCACTG GCAGCAGCCA CTGGTAACAG
 4261 GATTAGCAGA GCGAGGTATG TAGGCGGTGC TACAGAGTTC TTGAAGTGGT GGCCTAACTA
 4321 CGGCTACACT AGAAGGACAG TATTTGGTAT CTGCGCTCTG CTGAAGCCAG TTACCTTCGG
 4381 AAAAGAGTT GGTAGCTCTT GATCCGGCAA ACAAAACACC GCTGGTAGCG GTGGTTTTT
 4441 TGTTTGCAAG CAGCAGATTA CGCGCAGAAA AAAAGGATCT CAAGAAGATC CTTTGATCTT
 4501 TTCTACGGGG TCTGACGCTC AGTGGAACGA AAACTCACGT TAAGGGATTT TGGTCACTG
 4561 ATTATCAAAA AGGATCTCA CCTAGATCCT TTTAAATTAA AAATGAAGTT TTAAATCAAT
 4621 CTAAAGTATA TATGAGTAACTT CTTGGTCTGA CAGTTACCAA TGCTTAATCA GTGAGGCACC
 4681 TATCTCAGCG ATCTGTCTAT TTGCTTCATC CATAGTTGCC TGACTCCCCG TCGTGTAGAT
 4741 AACTACGATA CGGGAGGGCT TACCATCTGG CCCCAGTGT GCAATGATAC CGCAGACCC
 4801 ACGCTCACCG GCTCCAGATT TATCAGCAAT AAACCAAGCCA GCCGGAAAGGG CCGAGCGCAG
 4861 AAGTGGTCCT GCAACTTTAT CCGCCTCCAT CCAGTCTATT AATTGTTGCC GGGAAAGCTAG
 4921 AGTAAGTAGT TCGCCAGTTA ATAGTTGCC CAACGTTGTT GCCATTGCTA CAGGCATCGT
 4981 GGTGTCACGC TCGTCGTTG GTATGGCTTC ATTCAAGCTCC GGTTCCTAAC GATCAAGGCG
 5041 AGTTACATGA TCCCCCATGT TGTGCAAAAA AGCGGTTAGC TCCCTCGGTG CTCCGATCGT
 5101 TGTCAAGAGT AAGTTGCCCG CAGTGTATC ACTCATGGTT ATGGCAGCAC TGCATAATTG

Figure 20 (continued):

5161 TCTTACTGTC ATGCCATCCG TAAGATGCTT TTCTGTGACT GGTGAGTACT CAACCAAGTC
5221 ATTCTGAGAA TAGTGTATGC GGGCACCAGAG TTGCTCTTGC CCGGCGTCAA TACGGGATAA
5281 TACCGCGCCA CATAGCAGAA CTTTAAAAGT GCTCATCATT GGAAAACGTT CTTCGGGCG
5341 AAAACTCTCA AGGATCTTAC CGCTGTTGAG ATCCAGTTCG ATGTAACCCA CTCGTGCACC
5401 CAACTGATCT TCAGCATCTT TTACTTTCAC CAGCGTTTCT GGGTGAGCAA AAACAGGAAG
5461 GCAAAATGCC GCAAAAAAGG GAATAAGGGC GACACGGAAA TGTTGAATAC TCATACTCTT
5521 CCTTTTCAA TATTATTGAA GCATTTATCA GGGTTATTGT CTCATGAGCG GATACATATT
5581 TGAATGTATT TAGAAAAATA AACAAATAGG GGTTCGGCGC ACATTTCCCC GAAAAGTGC
5641 ACCTGACGTC TAAGAAACCA TTATTATCAT GACATTAACC TATAAAAATA GGCGTATCAC
5701 GAGGCCCTTT CGTCTCGCGC GTTTCGGTGA TGACGGTGA AACCTCTGAC ACATGCAGCT
5761 CCCGGAGACG GTCACAGCTT GTCTGTAAGC GGATGCCGGG AGCAGACAAG CCCGTCAGGG
5821 CGCGTCAGCG GGTGTTGGCG GGTGTCGGGG CTGGCTTAAC TATGCGGCAT CAGAGCAGAT
5881 TGTACTGAGA GTGCACCATATA TGCGGTGTGA AATACCGCAC AGATGCGTAA GGAGAAAATA
5941 CGCGCATCAGG CGCCATTCGC CATTCAAGGCT GCGCAACTGT TGGGAAGGGC GATCGGTGCG
6001 GGCCTCTTCG CTATTACGCC AGCTGGCGAA AGGGGGATGT GCTGCAAGGC GATTAAGTTG
6061 GGTAACGCCA GGGTTTCCC AGTCACGACG TTGTAAAACG ACGGCCAGTG CCAAGC

//

Figure 21: Nucleic acid sequence of the known segment of the R15/appa transgene used for the generation of transgenic mice (SEQ ID NO:5).

LOCUS R15/appa 3470 bp DNA **SYN** 15-APR-2000
DEFINITION R15/appa transgene with vector sequences removed.
ACCESSION R15/appa
REFERENCE 1 (bases 1 to 3470)
SOURCE synthetic construct.
ORGANISM synthetic construct
artificial sequence.
KEYWORDS salivary proline-rich protein, acid glucose-1-phosphatase; appA
gene; periplasmic phosphoanhydride phosphohydrolase; artificial
sequence;
AUTHORS Golovan, S., Forsberg, C.W., Phillips, J.
JOURNAL Unpublished.

DEFINITION Rat salivary proline-rich protein (RP15) gene.
ACCESSION M64793 M36414
VERSION M64793.1 GI:206711
SOURCE Rat (Sprague-Dawley) liver DNA.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata;
Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 1748)
AUTHORS Lin, H.H. and Ann, D.K.
TITLE Molecular characterization of rat multigene family
encoding proline-rich proteins
JOURNAL Genomics 10, 102-113 (1991)
MEDLINE 91257817
FEATURES Location/Qualifiers
source 1..1748
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/tissue_type="liver"
/tissue_lib="cosmid genomic library"
misc_feature 1802-1810
/function=" consensus sequence for initiation in
higher eukaryotes "

FEATURES Location/Qualifiers

DEFINITION E. coli periplasmic phosphoanhydride phosphohydrolase (appA)
gene,

ACCESSION M58708 L03370 L03371 L03372 L03373 L03374 L03375
VERSION M58708.1 GI:145283
SOURCE Escherichia coli DNA.
ORGANISM Escherichia coli
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.

REFERENCE 1 (bases 1811..3109)
AUTHORS Dassa, J., Marck, C. and Boquet, P.L.
TITLE The complete nucleotide sequence of the Escherichia coli gene appA
reveals significant homology between pH 2.5 acid phosphatase
and glucose-1-phosphatase

Figure 21 (continued):

JOURNAL J. Bacteriol. 172 (9), 5497-5500 (1990)
 MEDLINE 90368616

FEATURES Location/Qualifiers
 Source 1811..3109
 /organism="Escherichia coli"
 /db_xref="taxon:562"
 sig_peptide 1811..1876
 /gene="appA"
 CDS 1811..3109
 /gene="appA"
 /standard_name="acid phosphatase/phytase"
 /transl_table=11
 /product="periplasmic phosphoanhydride phosphohydrolase"
 /protein_id="AAA72086.1"
 /db_xref="GI:145285"

 /translation="MKAILIPFLSLLIPLTPQSAFAQSEPELKLESVVIVSRHGVRAP
 TKATQLMQDVTPDAWPTWPVKLGWLTPRGELIAYLGHYQRQRLVADGLLAKKGCPQS
 GOVAlIADVDERTRKTGEAFAAGLAPDCAITVHTQADTSSPDPLFNPLKTGVCQLDNA
 NVTDAILSRAGGSIAADFTGHRQTAFRELERVINFPQSNLCLKREKQDESCSLTQALPS
 ELKVSADNVSLTGAVSLASMLTEIFLLQQAQGMPEPGWGRITDSHQWNTLLSLHNAQF
 YLLQRTPEVARSRATPLLDLIKTAITPHPPQKQAYGVTLPTSVLFIAIGHDTNLANLGG
 ALELNWTLPGQPDNTPPGELVFERWRLSDNSQWIQVSLVQFTLQOMRDKTPSLNT
 PPGEVKLTLAGCEERNAQGMCSLAGFTQIVNEARIPACSL"
 mat_peptide 1877 3106
 /gene="appA"
 /product="periplasmic phosphoanhydride phosphohydrolase"

 mutation replace(1817..1819, "gcg changed to gcc")
 /gene="appA"
 /standard_name="A3 mutant"
 /note="created by site directed mutagenesis"
 /phenotype="silent mutation"
 mutation replace(3092..3094, "ccg changed to ccc")
 /gene="appA"
 /standard_name=" P428 mutant"
 /note="created by site directed mutagenesis"
 /phenotype=" silent mutation "
 mutation replace(3095..3097, "gcg changed to gct")
 /gene="appA"
 /standard_name=" A429 mutant"
 /note="created by site directed mutagenesis"
 /phenotype=" silent mutation "

 polyA_signal 3262..3457
 /note="SV40 signals"

 BASE COUNT 1065 a 721 c 735 g 949 t
 ORIGIN
 1 GGATCCCCTT TGCTATGTAG TTTTTAATGG AAATTACAAC CCATAGTGTG TTGATAAATA
 61 GAGAGTCCTG TTTGGTTTAA GCAACCTCTG TTTCTCATAA ACTCCATAAA AACAGGAATA
 121 CTCTTTGTTT CTAGCATAAC CAAAAGATTG AGTGAATTGA AAACAATGTT CCCTTAGAGT
 181 ATAGGGTCTAA TAACCCCGAA AATATTACCA TGATACTGAG CATTGTAAG TATCTCATAG
 241 CATGTAGTAT CCATAGTCCA TCAATGAGAG AGACATTAA CATGATTTC ATTAATCAGG

Figure 21 (continued):

301 TGGAAAAGAC ATGACAACAT TCACAGGCAC TGCACAGAAC ATAGTGGTCC ACCTTGCACA
 361 TATTTCACTA AACTAGGTTT ATCTATTTG TTGCTTCTC TAACATCTCT GCAATGAAGC
 421 AGGTCAACAG TGCCACATAT CCTTTACTTA ACCTAAGGAA CACAAAAAAT TTTCTACATA
 481 TATCCTGGTT AGAGAGTGT TAAAATAAGT TTTCCAAGAA TGAAAAGAA ATGTTCTGAC
 541 TTAACAATTA AGACAGTATT TATTAAAGC AAGAAATATG AGGCACACAA GAAAATATT
 601 TGGGAAGAAA CCATTTGGGTG AACAAATATT CAAATAAAA TAGACAAACA TAGTTAATTG
 661 TAAAACATAT GTTGACCAAG CCCTCTTTT CAATAGGCTT AATGTGAATA AAATGTTAAA
 721 GATTCTCTT GGGTGGCTGC AAATTGTCCA CGAATAAGAC AAAATATAAA AATAAGGACT
 781 GAGTCTCACA AAATGAAAAG GAAATATATT CAGAAAGAGA ATCTTGAGAG AATGTGTTGT
 841 CACAAATTAA AGAAAACCTG TGTTGAATGA CATCCTGAGG CCTGAGCTAT TACTGACATT
 901 TAAGATAAAAG GTAACTGTAT ACATTGTCC CATTGAGGGG ACAAGAAAGC TGCTCTCATG
 961 TTCAGCTCTA TAATTCTTGC CTTAAACAAC TTAAATAGAA TGATTTAAAA TATGGAGCTG
 1021 TCCATGGACC TTTGAAATAT AAAATAGTCA AGCAACTTAT CAAGGAATTA CAGATTCTT
 1081 GATACTAACCA CAGGTAAATC CCACACGTGT TTTGAGACTA CATTGCTGCG GATTTTATTG
 1141 ATGTAATAGG TCACATGTTT TTCGGGCCAA TGTTGCTGTT ATTGCGTTAC TTCAAGAGAA
 1201 TAGTGGCAAC TGATGCTATG TATTCTAGGG GTTTGAAGTG ATGTTTCATG ATTGAAATTT
 1261 GTAAAAGAAT AACATCATCA TTCTTAACAA TAGAACATAT AAAGTCACAC AGAAGTGACA
 1321 GTGTTTAAGC TGTACTATTG ATCAAAGAAA TTTATTACCT TCAGTTCAA TGGAAATAAT
 1381 TACTGATAAT ACAAACATGT GTGAACACAC ACTAACCTA TCCAAATGCA CAGTGATACA
 1441 CAGAAAATAT TAGCAAGTAG AATGCAATAT TTATATAACG ATTGTTATT TCAATCAATT
 1501 GTATGTATCA ATATATGGGC TATTCTTCA CACATGATT TATTCAAATT TACTCTAATC
 1561 ATTGTTGAAC CATTAGAAA AGGCATACTG GCAACTTTTC CTTACCTCAT CCAGCTGGC
 1621 AAAAGTCCC A GTGTGGAGTA AAGGATGCAA GATTCTCTGC TCTGTTAAGT ATAAAATAAT
 1681 AGTATGAATT CAAAGGTGCC ATTCTCTGC TTCTAGTTT AAAGGCAGTG CTTGCTCTT
 1741 CCAGCACAGA TCTGGATCTC GAGGAGCTTG GCGAGATTTT CAGGAGCTAA GGAAGCTAAA
 1801 AGCCGCCACC ATGAAAGCCA TCTTAATCCC ATTGTTATCT CTTCTGATTC CGTTAACCCCC
 1861 GCAATCTGCA TTGCTCAGA GTGAGCCGG A GCTGAAGCTG GAAAGTGTGG TGATTGTCAG
 1921 TCGTCATGGT GTGCGTGCTC CAACCAAGGC CACGCAACTG ATGCAGGATG TCACCCCAGA
 1981 CGCATGGCCA ACCTGGCCCG TAAAACCTGGG TTGGCTGACA CCGCGCGGTG GTGAGCTAAT
 2041 CGCCTATCTC GGACATTACC AACGCCAGCG TCTGGTAGCC GACGGATTGC TGGCGAAAAAA
 2101 GGGCTGCCCG CAGTCTGGTC AGGTCGCGAT TATTGCTGAT GTCGACGAGC GTACCCGTAA
 2161 AACAGGCGAA GCCTTCGCCG CCGGGCTGGC ACCTGACTGT GCAATAACCG TACATACCCA
 2221 GGCAGATAACG TCCAGTCCCG ATCCGTTATT TAATCCTCTA AAAACTGGCG TTTGCCAACT
 2281 GGATAACGCG AACGTGACTG AC CGATCCT CAGCAGGGCA GGAGGGTCAA TTGCTGACTT
 2341 TACCGGGCAT CGGCAAACGG CGTTTCGCGA ACTGGAACGG GTGCTTAATT TTCCGCAATC
 2401 AAAACTGTGC CTTAAACCGT AGAAACAGGA CGAAAGCTGT TCATTAACGC AGGCATTACC
 2461 ATCGGAACTC AAGGTGAGCG CGCACAATGT CTCATTAACC GGTGCGGTAA GCCTCGCATE
 2521 AATGCTGACG GAGATATTTC TCCTGCAACA AGCACAGGG A TGCCGGAGC CGGGGTGGGG
 2581 AAGGATCACC GATTACACCC AGTGGAACAC CTTGCTAAGT TTGCTATAACG CGCAATTTTA
 2641 TTTGCTACAA CGCACGCCAG AGGGTCCCG CAGCCGCGCC ACCCCGTTAT TAGATTGAT
 2701 CAAGACAGCG TTGACGCCCG ATCCACCGCA AAAACAGCG TATGGTGTGA CATTACCCAC
 2761 TTCAGTGCTG TTTATGCCG GACACGATAC TAATCTGGCA AATCTCGGC GCGCACTGG
 2821 GCTCAACTGG ACGCTTCCCG GTCAAGCCGA TAACACGCC CGAGGTGGTG AACTGGTGT
 2881 TGAACGCTGG CGTCGGCTAA GCGATAACAG CCAGTGGATT CAGGTTTCCG TGGTCTTCCA
 2941 GACTTTACAG CAGATGCCGTG ATAAAACGCC GCTGTCATTA AATACGCCGC CGGGAGAGGT
 3001 GAAACTGACC CTGGCAGGAT GTGAAGAGCG AAATGCCAG GGCATGTGTT CGTTGGCAGG
 3061 TTTTACGCAA ATCGTGAATG AAGCACGCAT ACCCGCTTGC AGTTTGTAAAG GTATAAGGCA
 3121 GTTATTGGTG CCCTTAAACG CCTGGTGTCA CGCCTGAATA AGTGATAATA AGCGGATGAA
 3181 TGGCAGAAAT TCGCCGGATC TTTGTGAAGG AACCTTACTT CTGTGGTGTG ACATAATTGG
 3241 ACAAAACTACC TACAGAGATT TAAAAAACCT CCCACACCTC CCCCTGAACC TGAAACATAA
 3301 AATGAATGCA ATTGTTGTTG TTAACTTGTT TATTGAGCT TATAATGGTT ACAAAATAAG
 3361 CAATAGCATC ACAAAATTCA CAAATAAAGC ATTTTTTCA CTGCATTCTA GTTGTGGTTT
 3421 GTCCAAACTC ATCAATGTAT CTTATCATGT CTGGATCGAT CCCCCGGGTAC

//

Figure 22: Nucleic acid sequence of the SV40/APPA+intron plasmid (SEQ ID NO:6).

LOCUS SV40/APPA 5421 bp DNA CIRCULAR SYN 14-APR-2000
 DEFINITION Ligation of SV40 promoter/enhancer into CAT/APPA+intron
 ACCESSION SV40/APPA
 REFERENCE 1 (bases 1 to 5421)
 SOURCE synthetic construct.
 ORGANISM synthetic construct
 artificial sequence.
 KEYWORDS SV40 promoter/enhancer, acid glucose-1-phosphatase; appA gene;
 periplasmic phosphoanhydride phosphohydrolase; artificial
 sequence;
 AUTHORS Golovan, S., Forsberg, C.W., Phillips, J.
 JOURNAL Unpublished.

DEFINITION *E. coli* periplasmic phosphoanhydride phosphohydrolase (appA) gene,

ACCESSION M58708 L03370 L03371 L03372 L03373 L03374 L03375
 VERSION M58708.1 GI:145283
 SOURCE Escherichia coli DNA.
 ORGANISM Escherichia coli
 Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 Escherichia.

REFERENCE 1 (bases 40 1337)
 AUTHORS Dassa, J., Marck, C. and Boquet, P.L.
 TITLE The complete nucleotide sequence of the Escherichia coli gene appA
 reveals significant homology between pH 2.5 acid phosphatase
 and glucose-1-phosphatase
 JOURNAL J. Bacteriol. 172 (9), 5497-5500 (1990)
 MEDLINE 90368616

FEATURES Location/Qualifiers
 Source 40 1337
 /organism="Escherichia coli"
 /db_xref="taxon:562"
 sig_peptide 40.. 105
 /gene="appA"
 CDS 40 1337
 /gene="appA"
 /standard_name="acid phosphatase/phytase"
 /transl_table=11
 /product="periplasmic phosphoanhydride phosphohydrolase"
 /protein_id="AAA72086.1"
 /db_xref="GI:145285"

 /translation="MKAILIPFLSLLIPLTPQSAFAQSEPELKLESVVIVSRHGVRAP
 TKATQLMQDVTPDAWPTWPVKGWLTPRGELIAYLGHYQRQRLVADGLLAKKGCPQS
 GQVAIIDVDERTRKTEGAAGLAPDCAITVHTQADTSSPDPLFNPLKTGVCQLDNA
 NVTDAILSRAGGSIADFTGHRQTAFRELERVLFNFPQSNLCLKREKQDESCSLTQALPS
 ELKVSADNVSLTGAVSLASMLTEIFLQQAQGMPEPGWGRITDSHQWNTLLSLHNAQF
 YLLQRTPEVARSRATPLLDIKTALTPHPPQKQAYGVTLPSTSVLFIAGHDTNLANLGG
 ALELNWTLPGQPDNTPPGELVFERWRRLSDNSQWIQVSLVFTQLQOMRDKTPSLNT
 PPGEVKLTLAGEERNAQGMCSLAGFTQIVNEARI PACSL"
 mat_peptide 106 1334
 /gene="appA"

Figure 22 (continued):

```

/product="periplasmic phosphoanhydride phosphohydrolase"

mutation      replace(46.. 48, "gcg changed to gcc")
/gene="appA"
/standard_name="A3 mutant"
/note="created by site directed mutagenesis"
/phenotype="silent mutation"
mutation      replace(1320..1322, "ccg changed to ccc")
/gene="appA"
/standard_name=" P428 mutant"
/note="created by site directed mutagenesis"
/phenotype=" silent mutation "
mutation      replace(1323..1325, " gcg changed to gct")
/gene="appA"
/standard_name=" A429 mutant"
/note="created by site directed mutagenesis"
/phenotype=" silent mutation "

DEFINITION Plasmid pBLCAT3 (bases 2200 to 4924)
ACCESSION X64409
VERSION X64409.1 GI:58163
SOURCE synthetic construct.
ORGANISM synthetic construct
artificial sequence.
REFERENCE 1 (bases 2200 to 4924)
AUTHORS Luckow, B.H.R.
TITLE Direct Submission
JOURNAL Submitted (06-FEB-1992) B.H.R. Luckow, German Cancer Res
Center, Im Neuenheimer Feld 280, W-6900 Heidelberg, FRG
REFERENCE 2 (bases 2200 to 4924)
AUTHORS Luckow, B. and Schutz, G.
TITLE CAT constructions with multiple unique restriction sites
for
the functional analysis of eukaryotic promoters and
regulatory
elements
JOURNAL Nucleic Acids Res. 15 (13), 5490 (1987)
MEDLINE 87260024
COMMENT Promoterless CAT vector for transient transfection
experiments
with eukaryotic cells. Allows the analysis of foreign
promoters and enhancers.
FEATURES Location/Qualifiers
source      2200 to 4924
/organism="synthetic construct"
/db_xref="taxon:32630"

SV40 t intron 1380..1993
/note="SV40 signals"
polyA_signal 1990..2230
/note="SV40 signals"
CDS          complement(3471..4317)
/codon_start=1
/transl_table=11
/gene="Amp"
/product="beta-lactamase"
/protein_id="CAA45753.1"
/db_xref="GI:58165"

```

Figure 22 (continued):

SV40 promoter/enhancer 5023..5402
 /note="SV40 signals"

BASE COUNT	1413	a	1321	c	1331	g	1355	t
ORIGIN								
1	CGAGATTTTC	AGGAGCTAAG	GAAGCTAAAA	GCGGCCACCA	TGAAAGCCAT	CTTAATCCCA		
61	TTTTTATCTC	TTCTGATTCC	GTAAACCCCG	CAATCTGCAT	TCGCTCAGAG	TGAGCCGGAG		
121	CTGAAGCTGG	AAAGTGTGGT	GATTGTCAGT	CGTCATGGTG	TGCGTGTCC	AACCAAGGCC		
181	ACGCAACTGA	TGCAGGATGT	CACCCCGAGAC	GCATGGCCAA	CCTGGCCGGT	AAAACGGGT		
241	TGGCTGACAC	CGCGNGGTGG	TGAGCTAATC	GCCTATCTCG	GACATTACCA	ACGCCAGCGT		
301	CTGGTAGCCG	ACGGATTGCT	GGCGAAAAG	GGCTGCCCGC	AGTCTGGTCA	GGTCGCGATT		
361	ATTGCTGATG	TCGACGAGCG	TACCCGTAAA	ACAGGGAAAG	CCTTCGCCGC	CGGGCTGGCA		
421	CCTGACTGTG	CAATAACCGT	ACATACCCAG	GCAGATACGT	CCAGTCCCCA	TCCGTTATTT		
481	AATCCTCTAA	AAACTGGCGT	TTGCCAACTG	GATAACCGA	ACGTGACTGA	CGCGATCCTC		
541	AGCAGGGCAG	GAGGGTCAAT	TGCTGACTTT	ACCGGGCATC	GGCAAACGGC	GTTTCGCGAA		
601	CTGGAACGGG	TGCTTAATT	TCCGCAATCA	AACTTGTGCC	TTAACAGTGA	GAAACAGGAC		
661	GAAAGCTGTT	CATTAACCGA	GGCATTACCA	TCGGAACCTA	AGGTGAGCGC	CGACAATGTC		
721	TCATTAACCG	GTGCGGTAAAG	CCTCGCATCA	ATGCTGACGG	AGATATTTCT	CCTGCAACAA		
781	GCACAGGGAA	TGCCGGAGCC	GGGGTGGGGA	AGGATCACCG	ATTACACCCA	GTGGAACACCC		
841	TTGCTAAGTT	TGCATAACCG	GCAATTAT	TTGCTACAAAC	GCACGCCAGA	GGTGGCCCGC		
901	AGCCGCGCCA	CCCCGTTATT	AGATTGATC	AAGACAGCGT	TGACGCCCA	CCACCGCAAA		
961	AACAGGCAGT	TGGTGTGACA	TTACCCACTT	CAGTGTGTT	TATGCCCGA	CACGATACTA		
1021	ATCTGGCAAA	TCTCGGGCGC	GCACCTGGAGC	TCAACTGGAC	GCTTCCCCTG	CAGCCGGATA		
1081	ACACGCCGCC	AGGTGGTGA	CTGGTGTGTTG	AACGCTGGCG	TGCGCTAACG	GATAACAGCC		
1141	AGTGGATTCA	GGTTTCGCTG	GTCTTCCAGA	CTTTACAGCA	GATGCGTGT	AAAACGCCG		
1201	TGTCATTAAA	TACGCCGCC	GGAGAGGGTGA	AACTGACCC	GGCAGGATGT	GAAGAGCGAA		
1261	ATGCGCAGGG	CATGTGTTCG	TTGGCAGGTT	TTACGCAAAT	CGTGAATGAA	GCACGCATAC		
1321	CCGCTTGCAG	TTTGTAAGGC	AGTTATTGCT	GCCCTTAAAC	GCCTGGTGT	ACGCCTGAAT		
1381	AAAGTATAAT	AAGCGGATGA	ATGGCAGAAA	TTCGCCGGAT	CTTGTGAAAG	GAACCTTACT		
1441	TCTGTGGTGT	GACATAATTG	GACAAACTAC	CTACAGAGAT	TTAAAGCTCT	AAGGTAATA		
1501	TAAAATTTT	AAAGTGTATAA	TGTGTTAAAC	TACTGATTCT	AATTGTTGT	GTATTITAGA		
1561	TTCCAACCTA	TGGAACGTGAT	GAATGGGAGC	AGTGGTGGAA	TGCCTTTAAT	GAGGAAAACC		
1621	TGTTTGCTC	AGAAGAAATG	CCATCTAGTG	ATGATGAGGC	TACTGCTGAC	TCTAACATT		
1681	CTACTCCTCC	AAAAAAGAAG	AGAAAGGTAG	AAGACCCAA	GAACCTTCT	TCAGAATTG		
1741	TAAGTTTTT	GAGTCATGCT	GTGTTAGTA	ATAGAACTCT	TGCTTGCTT	GCTATTACAA		
1801	CCACAAAGGA	AAAAGCTGCA	CTGCTATACA	AGAAAATTAT	GGAAAAATAT	TCTGTAACCT		
1861	TTATAAGTAG	GCATAACAGT	TATAATCATA	ACATACTGTT	TTTTCTTACT	CCACACAGGC		
1921	ATAGAGTGT	TGCTATTAAAT	AACTATGCTC	AAAAATTGTC	TACCTTTAGC	TTTTAATT		
1981	GTAAAGGGGT	TAATAAGGAA	TATTTGATGT	ATAGTGCCTT	GAATAGAGAT	CATAATCAGC		
2041	CATACCAACAT	TTGTAGAGGT	TTTACTTGCT	TTAAAAAAC	TCCCACACCT	CCCCCTGAAC		
2101	CTGAAACATA	AAATGAATGC	AATTGTTGTT	GTAAACTTGT	TTATTGCGAC	TTATAATGGT		
2161	TACAAAATAA	GCAATAGCAT	CACAAATTTC	ACAAATAAAG	CATTTTTTTC	ACTGCATTCT		
2221	AGTTGTGGTT	TGTCAAAC	CATCAATGTA	TCTTATCATG	TCTGGATCGA	TCCCCGGGTA		
2281	CCGAGCTCGA	ATTCGTAATC	ATGGTCATAG	CTGTTCTGT	TGTGAAATTG	TTATCCGCTC		
2341	ACAATTCCAC	ACAACATACG	AGCCGGAAAGC	ATAAAGTGT	AAGCCTGGGG	TGCTTAATG		
2401	GTGAGCTAAC	TCACATTAAT	TGCGTTGCGC	TCACTGCCG	CTTTCAGTC	GGGAAACCTG		
2461	TCGTGCCAGC	TGCATTAATG	AATCGCCAA	CGCGCGGGGA	GAGGCGGTTT	GCGTATTGGG		
2521	CGCTCTTCCG	CTTCCTCGCT	CACTGACTCG	CTGCGCTCGG	TCGTTCGCT	CGGGCGAGGC		
2581	GTATCAGCTC	ACTCAAAGGC	GGTAATACGG	TTATCCACAG	AATCAGGGGA	TAACCGAGGA		
2641	AAGAACATGT	GAGCAAAGG	CCAGCAAAG	GCCAGGAACC	GTAAAAAGGC	CGCGTTGCTG		
2701	GCGTTTTCC	ATAGGCTCCG	CCCCCTGAC	GAGCATCACA	AAAATCGACG	CTCAAGTCAG		
2761	AGGTGGCGAA	ACCCGACAGG	ACTATAAAGA	TACCAAGCGT	TTCCCCCTGG	AAGCTCCCTC		
2821	GTGCGCTCTC	CTGTTCCGAC	CCTGCCGCTT	ACCGGATACC	TGTCCGCCTT	TCTCCCTTCG		
2881	GGAAGCGTGG	CGCTTTCTCA	ATGCTCACCG	TGTAGGTATC	TCAGTTCGGT	GTAGGTCGTT		
2941	CGCTCCAAGC	TGGGCTGTGT	GCACGAACCC	CCCCTTCAGC	CCGACCGCTG	CGCCTTATCC		
3001	GGTAACATATC	GTCTTGAGTC	CAACCCGGTA	AGACACGACT	TATCGCCACT	GGCAGCAGCC		
3061	ACTGGTAACA	GGATTAGCAG	AGCGAGGTAT	GTAGGCCGTG	CTACAGAGTT	CTTGAAGTGG		

09/926375 DE22602

PCT/CA00/00430

WO 00/64247

Figure 22 (continued):

3121 TGGCCTAAGT ACGGCTACAC TAGAAGGACA GTATTTGGTA TCTGCGCTCT GCTGAAGCCA
3181 GTTACCTTCG GAAAAAGAGT TGGTAGCTCT TGATCCGGCA AACAAACCAC CGCTGGTAGC
3241 GGTGGTTTTT TTGTTTGCCTA GCAGCAGATT ACGCGCAGAA AAAAAGGATC TCAAGAAGAT
3301 CCTTTGATCT TTTCTACGGG GTCTGACGCT CAGTGGAAAGC AAAACTCAGC TTAAGGGATT
3361 TTGGTCATGA GATTATCAA AAGGATCTTC ACCTAGATCC TTTTAAATTAA AAAATGAAGT
3421 TTTAAATCAA TCTAAAGTAT ATATGAGTAA ACTTGGTCTG ACAGTTACCA ATGCTTAATC
3481 AGTGAGGCAC CTATCTCAGC GATCTGTCTA TTTCGTTCAT CCATAGTTGC CTGACTCCCC
3541 GTCGTGAGA TAACTACGAT ACAGGGAGGGC TTACCATCTG GCCCCAGTGC TGCAATGATA
3601 CCGCGAGACC CACGCTCACC GGCTCCAGAT TTATCAGCAA TAAACCAGCC AGCCGGAAGG
3661 GCCGAGCGCA GAAGTGGTCC TGCAACTTTA TCCGCCTCCA TCAGTCTAT TAATTGTTGC
3721 CGGGAAGCTA GAGTAAGTAG TTGGCCAGTT AATAGTTGC GCAACGTTGT TGCCATTGCT
3781 ACAGGCATCG TGGTGTCACT CTCGTCGTTT GGTATGGCTT CATTCACTC CGGTTCCCAA
3841 CGATCAAGGC GAGTTACATG ATCCCCCATG TTGTGCAAAA AAGCGGTTAG CTCCCTCGGT
3901 CCTCCGATCG TTGTCAGAAG TAAAGTTGGCC GCAGTGTAT CACTCATGGT TATGGCAGCA
3961 CTGCATAATT CTCTTACTGT CATGCCATCC GTAAGATGCT TTTCTGTGAC TGGTGAGTAC
4021 TCAACCAAGT CATTCTGAGA ATAGTGTATG CGGCGACCGA GTTGCCTTG CCCGGCGTCA
4081 ATACGGGATA ATACCGGCC ACATAGCAGA ACTTTAAAAG TGCTCATCAT TGGAAAACGT
4141 TCTTCGGGGC GAAAACCTTC AAGGATCTTA CCGCTGTTGA GATCCAGTTC GATGTAACCC
4201 ACTCGTGCAC CCAACTGATC TTCAGCATCT TTTACTTCA CCAGCGTTTC TGGGTGAGCA
4261 AAAACAGGAA GGCAAAATGC CGCAAAAAG GGAATAAGGG CGACACGGAA ATGTTGAATA
4321 CTCATACTCT TCCTTTTCA ATATTATTGA AGCATTATC AGGGTTATTG TCTCATGAGC
4381 GGATACATAT TTGAATGTAT TTAGAAAAAT AAACAAATAG GGGTTCCCGC CACATTCCCC
4441 CGAAAAGTGC CACCTGACGT CTAAGAAACC ATTATTATCA TGACATTAAC CTATAAAAAT
4501 AGGCGTATCA CGAGGCCCTT TCGTCTCGCG CGTTTCCGGTG ATGACCGGTGA AAACCTCTGA
4561 CACATGCAGC TCCCGGAGAC GGTACAGCT TGTCTGTAAG CGGATGCCGG GAGCAGACAA
4621 GCCCGTCAGG GCGCGTCAGC GGGTGTGGC GGGTGTGGG GCTGGCTTAA CTATCGGCA
4681 TCAGAGCAGA TTGTACTGAG AGTGCACCAT ATGCGGTGTG AAATACCGCA CAGATCGTA
4741 AGGAGAAAAT ACCGCATCAG GCGCCATTG CCAATTCAAGC TGCGCAACTG TTGGGAAGGG
4801 CGATCGGTGC GGGCCTCTTC GCTATTACGC CAGCTGGCGA AAGGGGGATG TGCTGCAAGG
4861 CGATTAAGTT GGGTAACGCC AGGGTTTCC CAGTCACGAC GTTGTAAAAG GACGGCCAGT
4921 GCCAAGCTTT ACACTTTATG CTTCCGGCTC GTATGTTGTG TGGAAATTGTG AGCGGATAAAC
4981 AATTCACAC AGGAAACAGC TATGACCATG ATTACGAATT CGGCGCAGCA CCATGGCCTG
5041 AAATAACCTC TGAAAGAGGA ACTTGGTTAG GTACCTTCTG AGGCGGAAAG AACCAAGCTGT
5101 GGAATGTGTG TCAGTTAGGG TGTGGAAAGT CCCCAGGCTC CCCAGCAGGC AGAAGTATGC
5161 AAAGCATGCA TCTCAATTAG TCAGCAACCA GGTGTGGAAA GTCCCCAGGC TCCCCAGCAG
5221 GCAGAAAGTAT GCAAAGCATG CATCTCAATT AGTCAGCAAC CATAGTCCCG CCCCTAACTC
5281 CGCCCATCCC GCCCCTAAGT CCGCCAGTT CCGCCCATTC TCCGCCCAT GGCTGACTAA
5341 TTTTTTTAT TTATGCAGAG GCCGAGGCCG CCTCGGCCTC TGAGCTATTG CAGAAGTAGT
5401 GAGGAGGCTC GAGGAGCTTG G

//

Figure 23. The nucleic acid sequence of the Lama2/APPA transgene used for the generation of transgenic mice and transgenic pigs (SEQ ID NO: 7)

LOCUS transgene 17732 bp DNA SYN 14-APR-2000
 DEFINITION Lama-appA cut XhoI..20623 to NotI..17732
 ACCESSION transgene
 KEYWORDS parotid secretory protein; acid glucose-1-phosphatase; appA gene; periplasmic phosphoanhydride phosphohydrolase; artificial sequence; cloning vector
 REFERENCE 1 (bases 1 to 17732)
 AUTHORS Golovan, S., Forsberg, C.W., Phillips, J.
 JOURNAL Unpublished.

FEATURES
 DEFINITION M. musculus Psp gene for parotid secretory protein.
 ACCESSION X68699
 VERSION X68699.1 GI:53809
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 3777 to 5332;)
 AUTHORS Svendsen, P., Laursen, J., Krogh-Pedersen, H. and Hjorth, J.P.
 TITLE Novel salivary gland specific binding elements located in the PSP proximal enhancer core
 JOURNAL Nucleic Acids Res. 26 (11), 2761-2770 (1998)
 MEDLINE 98256451
 REFERENCE 2 (bases 7147 to 12653; 13952 to 17731)
 AUTHORS Mikkelsen, T.R.
 TITLE Direct Submission
 JOURNAL Submitted (07-OCT-1992) T.R. Mikkelsen, Department of Molecular Biology, University of Aarhus, CF Mollers Alle 130, 8000 Aarhus, DENMARK
 REFERENCE 3 (bases 7147 to 12653; 13952 to 17731)
 AUTHORS Laursen J, Hjorth JP
 TITLE A cassette for high-level expression in the mouse salivary glands.
 JOURNAL Gene 1997 Oct 1;198(1-2):367-72
 MEDLINE 9370303

FEATURES Location/Qualifiers
 source 1..to 12653; 13952 to 17731
 /organism="Mus musculus"
 /strain="C3H/As"
 /db_xref="taxon:10090"
 /chromosome="2"
 /map="Estimate: 69 cM from centromere"
 /clone="Lambda YP1, Lambda YP3, Lambda YP7"
 /clone_lib="Lambda-PHAGE (Lambda L47.1)"
 /germline
 /note="Allele: b"
 misc_feature 3777-5332
 /gene="PSP"
 /function="salivary gland specific positive acting regulatory region"
 enhancer 7147..8724

Figure 23 (continued):

```

        /evidence=experimental
exon    11778..11824
        /gene="Psp"
        /note="exon a"
        /number=1
        /evidence=experimental
exon    12626.. 14190
        /gene="Psp"
        /note="exon b fused with exons h and i"
misc_feature 12644-12652
        /function=" consensus sequence for initiation in higher
eukaryotes ."
misc_feature 13952-13965
        /function=" M13mp18 polylinker"

```

DEFINITION *E. coli* periplasmic phosphoanhydride phosphohydrolase (appA) gene,

```

ACCESSION M58708 L03370 L03371 L03372 L03373 L03374 L03375
VERSION M58708.1 GI:145283
SOURCE Escherichia coli DNA.
ORGANISM Escherichia coli
Bacteria; Proteobacteria; gamma subdivision;
Enterobacteriaceae;
Escherichia.

```

```

REFERENCE 1 (bases 12653..13951)
AUTHORS Dassa,J., Marck,C. and Boquet,P.L.
TITLE The complete nucleotide sequence of the Escherichia coli
gene appA reveals significant homology between pH 2.5
acid phosphatase and glucose-1-phosphatase
JOURNAL J. Bacteriol. 172 (9), 5497-5500 (1990)
MEDLINE 90368616

```

FEATURES	Location/Qualifiers
Source	12653..13951 /organism="Escherichia coli" /db_xref="taxon:562"
sig_peptide	12653..12718
/gene="appA"	
CDS	12653 13951 /gene="appA" /standard_name="acid phosphatase/phytase" /transl_table=11 /product="periplasmic phosphoanhydride phosphohydrolase" /protein_id="AAA72086.1" /db_xref="GI:145285"

```

/translation="MKAILIPFLSLLIPLTPQSAFAQSEPELKLESVVIVSRHGVRAP
TKATQLMQDVTPTDAWPTWPVKLGWLTPRGGELIAYLGHYQRQLVADGLLAKKGCPQS
GQVAIIDVDERTRKGEAAGLAPDCAITVHTQADTSSPDPLFNPLKTGVCQLDNA
NVTDAILSRAGGSIADFTGHRQTAFLERVLNFPQSNLCLKREKQDESCSLTQALPS

```

Figure 23 (continued):

ELKVSADNVSLTGAVALASMLTEIFLLQQAQGMPEPGWGRITDSHQWNTLLSLHNAQF
 YLLQRTPEVARSRATPLLDLIKTALTPHPPQKQAYGVTLPSTVLFIAHGHDTNLANLGG
 ALELNWTLPQPDNTPPGELVFERWRRLSDNSQWIQVSLVFQTLQQMQRDKTPLSLNT
 PPGEVKLTLAGCEERNAQGMCSLAGFTQIVNEARIPACSL"

mat_peptide 12719 13948
 /gene="appA"
 /product="periplasmic phosphoanhydride phosphohydrolase"

mutation replace(12659.. 12661, "gcg changed to gcc")
 /gene="appA"
 /standard_name="A3 mutant"
 /note="created by site directed mutagenesis"
 /citation=[3]
 /phenotype="silent mutation"
 mutation replace(13934..13936, " ccg changed to ccc")
 /gene="appA"
 /standard_name=" P428 mutant"
 /note="created by site directed mutagenesis"
 /citation=[3]
 /phenotype=" silent mutation "
 mutation replace(13937..13939, " gcg changed to gct")
 /gene="appA"
 /standard_name=" A429 mutant"
 /note="created by site directed mutagenesis"
 /citation=[3]
 /phenotype=" silent mutation "

BASE COUNT 4719 a 4125 c 4168 g 4719 t
 ORIGIN

1 TCGAGAGTAT CTTTGTCAAGC TGTGCCTCCA ACAAAGGGGT ACTGTTGCC ACATAGAAAG
 61 ATCTAAACTA ATTAATTAAT CCCTCACCCG CAAATCTTC AGTCACTAAG TTAGCACGAT
 121 TGTGAACAA GTTCTCCAAA GGAGAGATAAC AGATGAGTGC GTATAGGGTG GACCTGGCTG
 181 CTGAGGGAGAC ACCTGCATCT GACTAAGAAG AGCCACGGTG TTAGTTGAAT GGTGTGGAGT
 241 AGGGTGGTTC TGTGGGACAG TAGAAAATCG AGAGGCATGT GCCGTTTAGT GAACTGATGG
 301 AAGCTACCCC AAACGACAGA GATTGTCAGT CAGGCCAATC CGTTTCGAGT TTGATGGGCA
 361 GCCGGACAGT GAGACAGACA CACCTACTCA GTTGGAGGAA GGATGAGAAC AATGGCCAGC
 421 AGGGATTGAG AGACCCTGAC AGGCGCAAGG CCCTAACACA CACACCTACC ACCTCACTTG
 481 ACAAAAGCTGC CAAAGACCAA AGACTTGTTC TCCATTAGAA ATGACAGCTG GCTTGACCCG
 541 ACAGCATAAT AAGCAGAGTG TACTCTGATT GGAGAACTTT AATGTGTTT ATTCACTGATT
 601 ATAAAAGGAC AGTATTACAG ATTTTGTGTT ACACTGCTGT TACATGTGGG GCAGTGTGTC
 661 TTTAAGTAGG GTAAAGTACT CTTTAAAAT GGGTCTCTAGA TATTTTTTCC TTTAACTCAA
 721 GTCTCTTACT GTTTAAATGA TTTTTATTTT GTTTAAATATG GAGGAAAAG AAGCGTAAAT
 781 GGACAATATA TATTTAGAGA AAGATGGTTA GCTGTCAGAA AAATATGCAA ATCAAAATCA
 841 CACCAAGACT GCAGCACACC CCTGTCAGAT GGCTGTGATC AAGAAAATAA ATGACAATGAA
 901 GTGGTGGTGA AGATGTACTA AAGGGAAACA CACACACACA CACACACACA
 961 CACACTGGAG CAACCACTGT GGAAATCAGT ATGAATGGTC CTCAAAACC TGAAGATAGA
 1021 CGGGGGCGTG GTGGCATACA CTTTTATTCC CAGCACTGGG GAGGCAGAGG CAGGTGGATC
 1081 TCTGAGTTCC AGGCCAGCCT GGTCTATAGC ACAGGTTCTA GGACAGCCAG GGCTACACAG
 1141 AAAAACCTG CTTGATTAA ACCAAACCAA ACCAAACCAA ACCAAACCAA ACCAAACCAA
 1201 ACCAAACCAA ACCAAACCAAG ACCAAACCAA AACACTGAAG ATAGAACTTC AGTATTCCAT
 1261 TCCTAGATAT ATACCCAAATG GAGACTAAGT CAGCAAGACA CCTGCACAGC CAGTGTTCACT
 1321 ACTACACTGT TCACCCACAGC CAGGCTGTGG AACCAGCCTG AGTGTCCATG ATAAATGAAT

Figure 23 (continued):

1381 GGATAGGTAA CTTTCAAGGT AAATGGACTC TGCTGTGTAC ATGCCTCACA TTCTGTTTAT
 1441 TCATTTTCTT TTATGAGGTG TCCATTCAAG AGTCACATGG TAGTTCTATT TTCAGTCTTC
 1501 TGAAGATACT ACACGGTCC CCACAGTTA CACTTTATC AGCAGTGAAT AAGGGTTCCT
 1561 CTATCCTTAC CATCATTGT TGTAATTTT CTTGATGACC CTCTTCTGA CAGGGATAGG
 1621 ATGTAATATC AGTGTGAGGA AGTACAACCT GTTTCTAAG TATTATTGG CCCCTGCAT
 1681 TTCTTCTTT GAAAAGTGTG GGTTCCTGAC ATCTGCTCAG GTATTCAATTG GATGTTGTTT
 1741 CTTGGTGTG TGAGTTCTA TGAATTCTAG ATGTTAAATC CCTGCCTGTG GTTCTCTCCC
 1801 ATTCTGTAGG CTGCCTCCCT ACCCTGGCAA TTGTTGTCCT GTTTTGCAG AAACTTTGA
 1861 CTTCATGGAA TCTCATTTGT CAGTTTCCC TCCTCTGCTA TAGCCTGAGC TAATGCACTG
 1921 GTTTTACAG AGCCCTGGTC TATGCCTTA TCCTCCTCTG GCAGCTTCGG AGTTICATT
 1981 CTTACATTTA GATCTTGAT CCACCTTGA CAAGTTTGG ACCAGGGTGA GAGATACGAA
 2041 TCTAGTTCCA TTCTTCCATA TGTGATCCTA GTTACATAG CATCGTTGGT TGAAGAGGTT
 2101 TTATTTATT TTTAAATAAT GTGTCAAAA AAACGAGGTG GTTGTAGCAG TGTGGATTG
 2161 TTTCTTGTG CTTGATCTA CAGGTCTTGT TTGTGTCAG TCTCATGATG TTTTATTGCT
 2221 ATGGCTCTGT CATACTGCT GAGGTCAAGG ATTGTGATAT ACCTTCAGTA TTGCTCCCTC
 2281 AGACTCAGGT TTGCTTGGC CAGGAGTCAT CTTACTCAGT GCTCTTAGAG CTCCCCCAGC
 2341 ATGTAGCTGC TACTATTCTT AGTTGATAAA TCAGGAAACT GGGGCTCAGA GAGATTAAC
 2401 GTCTTGAAC ACTCTCTGGG AGGTGAAAC TGGAGACACT AAACGTGTT TACCTGTAC
 2461 TGCTCCAGTA GCTGCTGGGT GCTGGGCTAC AGCAAAGCAC CTATACTATA TATTACTCAG
 2521 GAGGTGGAAA AACTCAGCCT CCCTTGGGGT TCCCAAGCTC CCAGGTGTCC AGTCACTGCT
 2581 GGAAACCTCA TGGAGTCTGA AAGGAAGGGT TGAGGGTACA TGGGGCAGGG ATGAGGAGCC
 2641 TGGGGCTGGG ATCTCCCAA CACCTGGATA TCCAGATGCC ACTGGGTCAAG GGGGAGTTGG
 2701 GAACAGAGTT GGGATGTCCA TGGACCTGTG ACAAGGCCAG GCCAGGGGG AGGATAAC
 2761 TGGCTTTACT AATTTCGAA AGTCCTTAGC TTAGCAGCAG TTGCTGGGA GCACAGAGGG
 2821 GCCTCTGTG AGAGGCTCAG GCAGTGCCTGC TCTGTAGGCG AAGGTCTTCT CCATGTTCCC
 2881 CATGGTGGTT CTTGATGAAA GAGACAGTCC TTGGCTCCAA ACTGGTTTAT TGATTGTTCA
 2941 TTGTGGAAAA TGGGTGCACA CCACCTTCTC AGGGTGGACC AGAGATCAA TACCTTTGC
 3001 AGGGAGGAAT ATCTGGGAAG GGACGCTTAC TGGCTAAACC CTCAGGGCCT CTAGATACAT
 3061 CATTAGCATG GAGAACTCTG TTCTGGCTA CATGACCACA GGCCACATTT CCACAAGCCA
 3121 CATGTGGAA GTGTGGCACA TGTTCTAGGC CAGGAATCTG GTAGGGAGCG TGGAGCCACC
 3181 TACCATCCCA GGTGGGTGCC TGGGTGCCAG GGACCTGAA CCCGCTCAAC CTTACCAAGT
 3241 TTCCTGGCAG GGTCCACTGT CCTACACAGA AGCTGGAGGA GGTGTGAGGG TTGTTCTT
 3301 GTGGAATGTC CCATGCTGCT TGGGGCTCAG TTTCTCCACC TGTACCTCAT TGGTTGGGT
 3361 ATAAAAAGTG GGGATACTTT ATTATTCTCT GACTCGTCC TGAGGAAAAA GCATCGTGGC
 3421 AGTCCAGGAA CCACACCCCTG AGGTTCTGC ACTGAAGGGA CTCCCTAAGT CTCTGGAGTC
 3481 TCTCCCTTC ACAGAGCTGC CAAAGTCTAG GTTCTTTGA GGATAACAGA GCCATGCTTG
 3541 GTAAGCAGAC AACAGCATT GTTACTCAA CCTTCTTTG TCAGCTCCCT CTTCATAAAC
 3601 AAGTTGAGAC ACCATGCTGG CTTGAGGAAG ACTTCTAAAG CCAGACAACT GTGCAAGGAA
 3661 GAAGAAGAAG GGGCAAGTGG AGTTAGCCTG GATGTAGCCC TCAAAGTCTC CAGAGACCAG
 3721 CCATGAAGGC TCAAGTGGAG GGCAAGACCT GCAGCAGCCA AGCATCTGGC AGGAGAGGAT
 3781 CCTGGGAACC CCTCTACCAT GACACACATT CTTCTGCTAG GTCACACTTA ATAGGCCATT
 3841 TCTTATTGGA ATCTATCATG GTGTTCTGTG CGAGATTAAT GAGGTGTAT GCTCGAACA
 3901 GAAAGTTATA TAAAAACAAG TCCCCCCCCC TTGTCACTGC TGCTAAGAAT GTAGCAGAAA
 3961 TTGTCCTCAAG TGTCTCTCTA ATCAGAAACA ATAAAGGTCT CTTGGATTC AAGCCCTCCA
 4021 GTTCTCTCCT TCCTTGCTGA GCCTTGGACA CCCATACAA CCTCCTGGAT GCTACAGCTC
 4081 TGGGCAGAGA CTCCAAGGTG GGGAGAGACT GATGGTACAA AAGCAAATA CTTGTTGGG
 4141 GGTACACCCA CTCTCTGCTC TGTGTGGTT CTGCACTCAG TCCCTGCAGAC AGGCCCTCAG
 4201 TGGGTCTTC ACAGAGCAACA CGCAGAGGGG GGCAATGGAT GGAATACCC ACACCCCTGGT
 4261 TAGTTTACCC CGGCCATGCT CTCTGCTCTT CATCCCTCCT CTGCCCTCTG CCACGGCTTT
 4321 CTCTGCAGGA ATCATATCTT CATATTGGCC CACAGGTGTT CTCCTCACCC TAGCTATGAT
 4381 GTTTACTTTA GAGTGACCTT AGCAGGGCTG GTGGGAATGA GTTCTAGAAG GCTCACGGAG
 4441 ATGCTAGGGA AGAAACGTCT TCTAACTACT GAGGTACTA AGTTCTGGT GTTGTCTCT
 4501 GCCTTCCCT TGTTAAAGTC ACCTTGAAGT TAGTGCAGAA GAAATCAGAG CCCAGTCACA
 4561 GAGTAAATAT GGTCTGAAG ATTTCTTTG AGTGCCTCAGA ATCCATGACA TTCAAGAGC
 4621 CCTCTTGTG CTTAAGTCA TTTGGGTTG TATCTCTGC TTGATGTATG TGTGTGTGTT
 4681 TATCAAAGAG TGAGATGGTT ACATAAGAGG TGCTCTAAAG GACAGAGAGG ATTTGCAATT
 4741 GTGGCATGTG ACATCCTCAG GCCTTGCTCT GGTGCCAGGA GGAACGTGATG CAGAAAAGAG
 4801 TAAGAGGTCA TTTCTGGAG GCTGTCACTA TAGAGGAGAT CTTACAGTGC ATTCCCTCCT

Figure 23 (continued):

4861 CCAGGCCCTG CCTGAGGATA GACATGTGCT GACTGCAACT GAAACAGAGG CTTGGGATGG
 4921 AGAGTTAGGT TCACAGAAGG GAGGGTGGGA GATGGATGCT TGCTGGGTC TGGTCTCAT
 4981 CACCAGCTCC TGACCACCCG GTCAGCCCAT GTGCTTATTC CATAGCTTC TTTGCTATG
 5041 TTTACTCAGT GTGGTGTTC TTGGGACCCA GCAGAAGCCA GTCCCAGGCT GACAGCTGTG
 5101 GATACACAGG GCAGCATGAG GGTCTCAGC CTGAAGCAGT CAGGCTGGCA GAAGAGAAAG
 5161 ACCAGCACAC ATTCTTCAA CCAACTATGT CTTGAAAAAC AAACATATTA TATCACATAT
 5221 ATTGCATTAA TGAGACAGCT AAAATGTACT CGGGTAGCAT GACTCCAGGT GGGGATATCT
 5281 GCAAGTGCCA TGAGTGGCAG AGGGACAGCC AATGTGAGGC AAGAAGGAAT TCTGGCTCAA
 5341 CACAGCTTAG CTCCCTGGT TTGGTTCAA CTTTGAGAGT TTGACCACAA GCACCTTATT
 5401 TTTGACATAT TTAAACAGAG CACAACTTG GGAAAAAGTT TTCTTATGAA AATTATCACA
 5461 ATAAAGCTTA AGGCATGACT ACATTTAAAT GCCTTGCAA AGTATATGTG CCCTCTTCA
 5521 CAAGAATGGT TCTATTGACT GAGAAATAAT GTTCAGGATA AAGATCCAGG AAGAAAAGAT
 5581 CAGGGATAAG TAAAATACTA AACTCTTTG CAAAGTACAT AGACCCCTTT TCATAACAAT
 5641 GGTTCTATT GACTGACAAG CACTGCTCAG GAGTTGGGAA AGAGTCTAGC ATAAGCACGA
 5701 TAGCCTGGAG ACTCTAGTGA GGCTCTAGTCT TACAGACAGC AAAAATCACC AGGTTACAAA
 5761 CTACATTCAAT TTCCAGTTT CTGATCAGGC ACAGGTATGA ATCCCTTCTG TTGAAGAGAA
 5821 AAGTCCATGT GTTTAAATAA TCTGGTTTCT CCAGTGTAT TAGCGAGAAG ACTTGAGCCC
 5881 TATACAACTC CCACCTGGAG TGACATCCTG TCTTCATGGT ATATTACATA CCTAGACACG
 5941 CTCATCTCAC AGACTTAGGA CTTTGTCTC TGATCTCCAT TTCTGATCCC ACTTCCACCT
 6001 TTGCTTGTGAT AGTGTCAATT TCTTCACTGC CTTGGTGACA ACCATGTAT CCTCTGTGTA
 6061 TTTGAGTGT ACCATTTCA GATTTTACCT GTATGCAAGA TCACACAGTC TTTGCTTT
 6121 TGTCTGGATG CATGCTAATC TCTACACAAAC AACCTTCCC CGTCACTCAG ATCTCCTCC
 6181 ATTAACACAT ACATGGTGC GAAGAGGCTA GGGAGCTTCC TTTCAGTGGG GAGCTAGCTG
 6241 GCTATTGGGC CTTTTGACT GTCCAGGAAG GCCCCCAATT GCTGAGACAA GAACCTAGAT
 6301 TCTTCATTAT TGACTCTAAC TCATGTATCA AGCAGAAGCT AATGAATAGT TATCAACAGG
 6361 ATCAGAGGTT CCAGTGTAAAG ACACTTTGAC ATGAAAAGAAC GGAGGAAGGA CAGATGGATG
 6421 CATAAAAGCA GGACCACTGC CCCAGGAAGG TCCTGGAAAC TGATGCAGGG CAAAGGACAG
 6481 GTTATAAACCC AAATCTTAGG GAGTCAGGAA GAGCACAGAG GAGCTCAACC AACTGACCAC
 6541 TGCTTAGGGG CTACCAACCC AATCCTCCCT GTGGGAACAG CTAAGCTATC AGCCAAGGGT
 6601 AATAAACAGG CAGGACCTGT GGATGACATG GAGAGCATAG GGACCCCTGGG TCCAGCCTT
 6661 AGCACCTGCA CTCTCAGGAT ACTCCACCAT TGTGCTTCTAG AGAGCCTAGG GATACTGGGT
 6721 CCAGCCTTTG GTACCTTCAC TCTCAGGGTA CCCCCTCACT GTGTCTTGGA GAGCCTAGGC
 6781 ACCCTGGTC CAGCCTTCAG TACCTGCGCT CTCAGGACAC CCCACCATTG TCTCTTGC
 6841 CGTCTCTTCT TCCTCTTCTC CCCCTTCACT GTCTCTTCTC TGTTTCTTTC TTGACTCTC
 6901 TTTCCCCTCA CACCCCTCACT CTAGTTCTC CCTTCCCTCT CTGCATCACC CTATTCTCTC
 6961 TGTGGTCCCT CCACTTTCTC TTATCTCTCA TGCTCTCTC CTCCTCTCAA TACTTGTCA
 7021 CCACTATACT TCAGGGGCCA GCTCTAGTGA CAAAGCTGTT AATAGCAAGA CTCTCAGATC
 7081 TCCAACGGCT CAGAGGAGCC AGACCCACCA AGAACCTCTC CCAGGTCCAA TTTCAGGTT
 7141 CTTCGAAAGC TTTCAGCAAA TGCTCAGGGG ACATGCCACT ACAAGAAGA TGCAAATTCC
 7201 AGTTGAGAGT GGGAAAGGCC CTTGCGTAGG TCCCCTCTC CAGGCCAAGG TCAGAGGGGC
 7261 TCTGTGTAAT CGGGATTGAC AGGGCTCAGA ACAATGTTT GTTTTAAGG TTATTTTATT
 7321 TTAGGTGTTA GTGTCTTGC TTGCATGACC TTATGTGCAT CATGTGTG CAGGTTCTG
 7381 ATGACAGTAG AGGAGGGCTT TGAATCCCTG GGGATAGGAA GTTACAGGAA ATTATAAGCT
 7441 GCTTGTGGG TCTTCTAGCT TTCCCAACAG AAGTGAATGC TCTTCACAC TGAGCCATCT
 7501 CTCTAGGCC AAGAGACATT GCTTTATGGA TATAATTGTG TGTGTGTC AACATTGAGG
 7561 AAAGGGAAAT AAAAAAAA CTTCAGCCGC TAAGGTTGTA CAGTTTCACT AATTGCTACT
 7621 TTAGTTGTG ATAAAATGGC AGGTGCTTC ACATTATAT ATACAAAAAC TTCCCTGCTG
 7681 GTGGTCAAC TGTGAGAACT GGGGTAAGTG GGTGAGTTCT CTTTTCTGT CTCTGTCT
 7741 GTCTCTCTCC TTCCATTCTT TCTTAAAGGA AATAAACATT GCAGCTGGGT TATAGCTCAT
 7801 CAATATGGAA GTTACAGAAG TGAAAAAAGG CATTGCCTG GTGGGTGGTG TTACAGCTG
 7861 ATTTTGGTT GTCCTGCAAG GAGGTCTGGG GACTGGCTGC TCTGTCTCTG TCTGTATGAG
 7921 TGAGGGAAAGT CTGGGGAGCA GATTCCCTAA CCTTCAGCCT GGCCTGGTCT CTGAGTGAAC
 7981 CCAGCCTCTC TGGCTCTAGT AGCTTTTCC AAACAGGAAT CTGAGTGGTG ACAGGGAACA
 8041 AGTACCAAGCC CATTGCTTAA GTGCCAGGGT TAGTGAGGGC AGGAAGCTGC CATAGCTGGG
 8101 ATTAGTAGTT GTATTGGATG TAGGAAGTCC TATCCTGGGA CAGCTAATCC TTAATGCTTC
 8161 ACTGGAGATT TTCAATGAGA AATTTATCCC ACGGCCATA TGGCCCCATC CTTTTGTCTC
 8221 CAACAGCCAA GTATTTCCA TTAGAGGAGA CTTCTGTAC ACTTGATGGA TGCTCATTCC
 8281 AAGGTGACTT GGGCAGTCA GTACAGACTT GGGATGACCT CTGACAGGCT AACCTCTCCC

Figure 23 (continued):

8341 CAACAAGGGC CCTCTATGTT TGCTATGTA TGTAATGTCA GACATTGTCA GGAGTGTCCG
 8401 CAGCACAGCC TGCCCAGTGT GAGGGCTCTC ATAGGTTCC CACTGTCTTA TCTACACAGG
 8461 GATAACGAGG AGGTAAAGCTG CAGTTCCCAG TCTCACTTCA CAGAGGAAGA GATAACCCCA
 8521 TCCCAGGTCA TGTAGCCAGC AGTGGAAAGA ATGAGGATTG GAACTCAGGT CTTCCAAGTC
 8581 CCATTGATAG CATCTCCTCA CAAGTCCCTT GCCACCCCTCA CGATGCCTTA GACACTTGCC
 8641 TGCCCTTTAT ACTAAGGAGA TGCAAGGTACA AGGGGTTTAC CCATGTAGCA GCTGAGGCAG
 8701 CTGGGGATAG ATACCAGCAG CAGGCCGTGAT GTCACCACTC TAACTCCAGC ATCCCCAGTC
 8761 TGTGTTCTG GAGTGTGAAA ATCCCTACTT ACAAGAGATTG TGCAACAGTC CTTGGCTCTG
 8821 TGACCCATAG CTGGAAACAG GATTCTCATT GATTGTGGA ACATGGTGGC AGCCAGCAA
 8881 AAAGAGGGTC TGCATACAGA AGACACGTGT GGCAAGGCCA CAGCAGACTC TGACTACCTT
 8941 AGCTTACAGA ATTACAAGGT CATAATGTCC TCTGCTTGG TCACCTCATG TTAAGGACAG
 9001 GCCCTAATGA AGATGGGGCA GAAGACTGAA GGAATGGCCA ACCAATAACT GGCCCAACTT
 9061 GAGACCCATC CTACAGGCAA GCATCAATTG CTGACACTAC TAATGATACT CTGTTATGCT
 9121 TGCAGACAGA AGCCTAGCAT AACTATCCTC CGAGAGGTCC ACCCAGCAAC TGACTGAAAC
 9181 AGAAAAAGAT ATCCACAGGC AAACAGTGGA TGGAGGTCA GGACTATTAT GGGAGAGCTG
 9241 TGGGAAGGAT TAAAAAACCTT GAAGGGGATA GGAACCCAC AGGAAGACCA ACAGAGTCAA
 9301 CTAAGAGACC TGTGGGAGCT CTCAGAGACT GAGCCACCAA CCAAAGAGCA TACACAGGCC
 9361 GGTCCGAGGC ACCTGGCAGC TGTAAGGCAG ACATGCAGCT CAGTCTCCAT GTAGGTCCTC
 9421 CAATAAGCGG TAGCCTGACT GCAGTATCCA ATCCCCAACAA GGGCTGCATA GTCTGGCTC
 9481 AGTGGGGGAG GATGCCCCCTA ATCCTGCAGA GACTTGATGA GTGGAGAGCT ATCCAGGGGG
 9541 AACCCACCCCT CTCTGAGAAG GGAATGGGGAA TGGGGGAGGG ACTCTGTGAA GAGGGGACAA
 9601 GGACAAACAA AAACCTCAAA TAGGTCAGGC CCTAAAGGCT TGCTAAGTAG CAGTGGCCCA
 9661 GCTCTGCTCT GTTCCTCAGC CCAAGGCTCA GCTCCCACCT GTTCTGTGT TTTCTGGCT
 9721 TTTCATGGGC CTAGGACTTG GTGACCAGTT CAAACAATGG GGCCTGTGGA AGACACAATA
 9781 TACAAGACTA GGGACATTTC TGTTCTGCTG ACTATCCATA GCCTGATGTA GGTGGAAGGA
 9841 CCCAATCACT GGATTCTAC CCTTGACAA CCTTGACAGC TGAGGGCCTC TCAGAAACCT
 9901 ATTTCTTCCA CTGAAAAATG AGACTCTAA ATGAACGTG TGACAATCAT CAGGTTTATT
 9961 AAAGAGGTGT ATCTAACCTG AATGGCAAGC AGACAGCAGG CAAATGTCTG TATCAACCTC
 10021 TAGGAAGGAC AAGAACTGCT CACTGCTGCC CCCCAGGAGG CCATTGCTG AAACAGCTGC
 10081 TCTCCTGCTG GTGCACAGGC CCTGCCTTCT CATTGCAAGCC ACAGCCCCCTT CCTGCTGAA
 10141 CCTCCTGTCA GGTCACTGGG AAACAGATCA AGATGGAACA GGACAGCTCC TGATGGTAA
 10201 TAAAAAACAG TGGTCATGGC TATTCTAGG GTTTATGCT TCTTCAGTCC ACACTGTGAA
 10261 GAGCTGTGGG CATGAACACC AGTGTTCGAG GTAGAGTTGG GTTCTGAAA TTCACAGTGG
 10321 GGTGAGCTCA GTAAATGTGA GCTGGAGGTG ACTCGTGAGA CACACAGTCC TGCTGCTTCT
 10381 GTTCCAATA TCCTGAGGAG ACGACACATC TACTTTGTTT ACAGGCCACA GTCTAGTTGA
 10441 CCTGAGAGTT ACCAGTTCT TATTGTGTG TGTGTGTGTG TGTGTGTGTG
 10501 TGTTGTCGT GTGTGAGTGC AGGTGCACAT ATGATAGCGT ACACGTTGAG GTCAGAGGAT
 10561 AACTATCAGG CGTTGTCCCC TCCTACTTTT CCTCGGACTC TGGAGAACAA ACATGGGTCC
 10621 TTATTCCAGG GGAGCAAGTC GCTGTTGGCT GACACATCTT GCTCACATAC ATTTACCTA
 10681 GACAATGGAG CCTCCATCAG AGTATTACTT TAGCTCCTCA CCGATGGCAA TGCACCCACCT
 10741 CTCTACCCAC ATAGGAGTT GGCTCTCCACA CACCCCCACA CCCCCCTTCAC CAAAACGTT
 10801 TCAGTTACTT TATCTGGTAA AGTCATCAG AGAATGAAGC CAGTATTAAG AACATGGAAT
 10861 CATTGGAA CCTGGATCTA GCAATACCCC ACCCTAGATG GAGTTGCTGA GTTTCACCT
 10921 CAGATTATAA TTCCCCCTTA GCTTCTATGG TTTATTCTGA AACCAGGGGA ACTCGATTCC
 10981 TCCCTTGGAA CCACAGACAT CCTGGCTTGT GAATTCACAT GTCATCTACT GCTAATCCAT
 11041 TGGTAGTATG TGGCTCACAG AGACACACTA CAGTCATGGC CAATGTCAAG GTAGGACAGA
 11101 TGTGAATCAT TCCCCCAGTC CTGCTGTTT CATGACTAAC CCTCCTCAGC ACAGTGACCA
 11161 TGAACCTACT TTTCCCCCTC TTTTATTTTT AGAATTGCTG GAATTTTCTA TTTTGAGAAA
 11221 TAATAGCCTT GGGCAGCATT AAACAAAATC ATCTAGAAAG CTGGTTAAA ATACAGATGG
 11281 TTGAGTCAGT GAAAGAGTGA GGAATGTCA TATTGGCCCC TCACAGAGGC TGGCTCACTC
 11341 CAGCAGAGGT GGTTGAAGCT CTTGGACACG GGTCAAGGTGC ATAGGAAAGG TNGTCTGGGA
 11401 CACTGAGAAC CACAATTGAA CAAACAGAAC TGTTGGCTT TTTTTTTTA AATGAGTTCT
 11461 CAAAAAAATGA CTGGCTAGCT TAGGCAAATA CTTCGAGCCA ACCCAACAGA ACATTCTCC
 11521 ATTGATTCACT TCTGGATCTT CTTCTAGAC AATACTGAAC TGACCCCTTG TTGGCAGTCT
 11581 CAAGTTGAC AACATAGGGC TTTGAACCTTG GCACAAGGTC CATCACTGTC ACCCAAGCAT
 11641 CCTGGGTGAC CTTTGGGTTG GAATATCTTG GCTAACCTTA GATATTCTT TTGGAGTATC
 11701 TTTAGAACAT CCAGGAAATA GGGCTTGATT CTCATCCTGG GACCACAATA TAAGTCACCC
 11761 TAGAATCCCA GGAGATCGTG CAGAGAAACA AGGATCTCTC TCGTGTGCA~~C~~ CTTCTTCAA

Figure 23 (continued):

11821 AGCAGTGAGT AGTGACTCCA CTAACACTGAG TTCCCACATCTG AGAGTCCACA GGAGGCTTTG
 11881 GGGCAAGAAG CAGAGGGAAAG GCACTGTTTG TGTTGGTAAA GTTTGACTC TAACAAATTT
 11941 GAAGACATAG ATGACATTGT GTCAGACTAA CAACAAACCTA GACTCATGTG GGTTCTGTG
 12001 AGGGATCAGA TTTTATTCAAT CAATGACTTG TCTTAGTGTAGAGAAAGG CTTCTACTG
 12061 GAGTGTAGGC TCAATAATGA CAGAAGAGAT AGCTATTTC CCTAGGGACT GTGCTGCTCC
 12121 AAGTTGGTG GAGAAAGGCA GTGGGGAAACC TAGATGTGCT CTCTGGGGAG GGGGCTGAA
 12181 GCTGGCTTCA TAGAAGGGTGT GAAGTTTGC TGAAACATCT AAACAGAATT ATAGCTTAGG
 12241 AAAGTGAGCA GGCAAGGCAG GGAATGTGTT GCATATGTAT ATGTACATGA ATATATTATG
 12301 TTATAGATAC ACACACATTT GAACCTCATT TGCAAGATGAC AGAAAATAGG TTATTTGCC
 12361 TCTCTTAACG GCTAACGACA ATGACTTCCA GTTCCATCCA TTTCCTGAAA TGCCACAATT
 12421 TCATTTTCA TTGTGGCTGA ATAAAATTCC ATTGCAGACT GGGCCCTACT TCATCCACTC
 12481 CTGAGGGCAG GCATATCCCC TGGCTCCATT TCTTACCTAT TGTGAAGAGA AGTGAACCTG
 12541 TCTTGTGAA AGGCAAGCGT GAGAGAGGCA GGCACTAATT GTGGGTTTT GTTTCTCTT
 12601 CCTGCTATGA CTCTCCATT GTCAAGAACCA AAGATCGATA AAAGCCGCCA CCATGAAAGC
 12661 CATCTTAATC CCATTTTAT CTCTCTGTAT TCCGTTAACCC CGCAATCTG CATTGCTCA
 12721 GAGTGAGCCG GAGCTGAAGC TGGAAAAGTGT GGTGATTGTC AGTCGTATG GTGTGCGTGC
 12781 TCCAACCAAG GCCACGCAAC TGATGCAGGA TGTCACCCCA GACGCATGGC CAACCTGGCC
 12841 GGTAAAACGT GGTTGGCTGA CACCGCGCGG TGGTGAGCTA ATCGCCTATC TCGGACATTA
 12901 CCAACGCCAG CGTCTGGTAG CCGACGGATT GCTGGCGAAAGGGCTGCC CGCAGTCTGG
 12961 TCAGGTGCG ATTATTGCTG ATGTCGACGA GCGTACCCGT AAAACAGGGG AAGCCTTCGC
 13021 CGCCGGGCTG GCACCTGACT GTGCAATAAC CGTACATACC CAGGCAGATA CGTCCAGTCC
 13081 CGATCCGTTA TTTAATCCCT TAAAAACTGG CGTTTGCCTA CTGGATAACG CGAACGTGAC
 13141 TGACGCGATC CTCAGCAGGG CAGGAGGGTC AATTGCTGAC TTTACGGGC ATCGCAAAC
 13201 GGCGTTTCGC GAACTGGAAC GGGTGCTTAA TTTTCCGCAA TCAAACTTGT GCCTTAAACG
 13261 TGAGAACACAG GACGAAAGCT GTTCATTAAC GCAGGCATTA CCATCGGAAC TCAAGGTGAG
 13321 CGCCGACAAT GTCTCATTAA CGGGTGCCTG AAGCCTCGCA TCAATGCTGA CGGAGATATT
 13381 TCTCCTGCAA CAAGCACAGG GAATGCCGGA GCCGGGGTGG GGAAGGATCA CCGATTACACA
 13441 CCAGTGGAAC ACCTTGCTAA GTTTCATTAAC CGCGCAATT TATTGCTAC AACGACGCC
 13501 AGAGGTTGCC CGCAGCCCGC CCACCCCGTT ATTAGATTG ATCAAGACAG CGTTGACGCC
 13561 CCATCCACCG CAAAAACAGG CGTATGGTGT GACATTACCC ACTTCAGTGC TGTTTATCGC
 13621 CGGACACGAT ACTAATCTGG CAAATCTCGG CGGCGCACTG GAGCTCAACT GGACGCTTCC
 13681 CGGTCAAGCCG GATAACACGC CGCCAGGTGG TGAACGGTG TTTGAACGCT GGCGTCGGCT
 13741 AAGCGATAAC AGCCAGTGGA TTCAGGTTTC GCTGGTCTTC CAGACTTTAC AGCAGATGCG
 13801 TGATAAAACCG CGGCTGTAT TAAATACGCC GCCCGGAGAG GTGAAACTGA CCCTGGCAGG
 13861 ATGTGAAGAG CGAAATGCGC AGGGCATGTG TTCGTTGGCA GTTTTACGC AAATCGTGA
 13921 TGAAGCACGC ATACCCGCTT GCAGTTGTA AGGTACCCGG GGATCACAAAC TTGCCCTCTG
 13981 AAGAGGAAGA ACAGAAGGGAT GCCACAACTC TCCTGCTGGC TACTCTCCAG TGGTTTCATC
 14041 TTACTTCTGA TGGCATTTCCTCTAGAAAG TGCTACTATC ATCCACACAT TTCTACCTGA
 14101 GACCACCCAA AGGACCCCTCC CAAATTCTCT TCCCTCTCTGA GTAGTCTCCA CACCTGTTAC
 14161 CACCATCCCA GAATTAAAAT CCTAACTGCA CTCTGGCGTG TGACTTGCCT CAGTCCTTGC
 14221 AATAAGAGTT GTTGGCAGTG CCAGGCCTGG TGCGCACGC CTTTAATTCC AGCACTTGGG
 14281 AGGCAGAGGC AGGGGGATTCTGAGTT CTGAGTTCGA GGCCAGCCTG GTCTACAGAG TGAGTCCAG
 14341 GACAGCCAGG GCTATACAGA GAAACCTGT GTCGAAAAAC CAAAAAAA AAAAAAAAGTT
 14401 GTTGGCAGAG TGTGGTTAT ATACCAGGTG GAGATTCAA ATGAGTGGCT GAAGCTGTAG
 14461 CCAGAAGGAA CTTAGAGGGAT AGCTCATAAC TTAAAAAGAA ATGTAGAGAG TAGCAGAAC
 14521 ATTGAGAGAG TGGGCACACCA GCCACTGTGT GAATGTGGCA GAACACAATC CAGCCAGCTA
 14581 TACATGCATA AGTGTATATT GGGGCCATCC TGACTGATGA GACACAGGAA AACAGATAGA
 14641 CGGGGTTAGG TGGCCATGGC CTTTCCCTGCC TGCCCTTCC TAAGGGTCTAT CTCAAGACCT
 14701 TATGCTCTCT TAACCTTCC ATTGCTACTT AGCTTCTAGA TATCACCTCC AGATTAGTCT
 14761 CCTTGGGTAC ATCAGTGATC CTGGTGTAT CCAGGCTTC CTGATTCCAT CTTTGTCTA
 14821 GAGGCTGCAA CTAAAGAGGT CTTCTTAATA CTTCACACCC TGATGCCAAA AGGAAGACAC
 14881 AGAAGTTCAC AGAGGTGAAG TGATTCAATG AGGACATACA GTGAGCAAGC ATCAGGGTCC
 14941 GGATTATCTG ACTCTACTCT AACCTTTATG TAAATGTGCT TTATGCCATT AACACTGTCA
 15001 TTCTGTGCT TCAGCTCTGG GAGACTCCCA AGCACTCTTA GGCACAAAGCC ACAATTAAGG
 15061 GACTCTGACA CTCTGCATTG ATTAATTAGC ATGGTGGTCT CTATGTTTCC AGATTCTG
 15121 TTGTTCACT TTCCATATAG GCTATGAAGG GTGTGAGGAA ATTTTTGGG GACAGAATTG
 15181 GAGGCAATCC ACCTCTCTCA GGAAGCCTCT ATCTGGAAAA GCTTACAATC CAGGGACAGT
 15241 AACTGTAGGC CCAGTCCTTG GTGTCCAAA TGGGTTTAT GTTTGAATC TGCAAGCCT

Figure 23 (continued):

15301 TCCATGTGCT CAAAGGTTG AACATGGAGC CTCCTCTGG TAACACTGTA TTGGAGGCTT
 15361 TTGAGACTGG ATGCTCTTG GTCCCATGTT TTGCTACATC ATCTGTCAAG ATATGACCA
 15421 GGCATGCTAC CAGCTACAC AGACTATGCC TCTCCAGCT TCATGTTCTC CCCACCATGA
 15481 TAGACTTGTA TCTCCTAAAA ATGGAATCAA AGCAAACCTT TCCTGCATTA AGTTTTTTT
 15541 TTTCTGTTAA GTGTTGGTC ACAGGGACAA GAAAACACTC AATACAGATA ATTAGTACCA
 15601 GAGTTGAGGT TCATTGCTCT AGCAAGTTG ATCAAATTT TAGGGCTTTG GAACTGATTT
 15661 ATAAGAGACA TGTAGAAGAG TCTGAAGCTG TGGGCTACAG AAGTGTCAAC AGTTTTTAAG
 15721 AATAGTTAA TACACCATGG GAATTGTGAA AATCAGAATG CTCACACAAA GGCAGACAGG
 15781 AAAACGTGAG CATGTGGCGT GTGAGAGGGC ATAAGAAGGA ACCTAGGGGG AAATGAGCTA
 15841 GAAGCCATTG GGCTACGTTA GGGAACGTGT GTGGCTGTG TTGGCCCATG CCCTGGCAAT
 15901 CTGAATGAGG CCAAATTTA AAGGAGTGG A CTAACCTCGAT TGTCAAGAGAA AATATCAAGA
 15961 CAGACCACCA CTCAGGCTAT GCCGTGTTG TGACCGACCA GCTACTCTT GCCAGCTCTA
 16021 TTGTGAAATT CCAGAGCAAT TATCAGAGCA TGAAGATACA TACAGTTTAG TGAAGTAAGG
 16081 GGTGTGGGTC CCTAAGTGGT TGGTGCATAA ATCTATGTAG GTGATGCCTA AGTGACACTT
 16141 GATAATCCAA AATATCAGCA ATGTGGAATG TCTTCCAAGG AGACCTGTAG ACACACATTT
 16201 TAGAACTTTG CTCATGGCTG TAATAAAATAG CTAGCTAGAA ATCATTTCCT GAAGAGGTTA
 16261 GTCTGAGTTA CGGTTCCAGG GCAAACATTC AGTGTGGCA AGGAAGGCAT TGCAGTCAGG
 16321 AGCCAAAGGT CAGCTGGTCA CATTGCATCA AGAGTAGAGA GTCAAGAGTGT GAGTAGAAAG
 16381 AGGATACAGG TTATAAAACC TCACTGTCCA CTCTCAGCAA TCCATTTCCT CCTAAAAGGC
 16441 TTTACCTTCT AAAGATTTTA GTCTTCAAA CCAGTACCAAG TAGCCTGGGA ACAAAAGTTG
 16501 AAACAAATGA GCCTTTGTGG GGCATTTCAC ACTTAAACAA GGGCATCACC TAGGAGGAGC
 16561 CCTGTGTGCA GTAGGAAGTG TGGCTCTGT GTCAGGAATG CTCAGGCTAA TAAGGGGTCC
 16621 TCTATCTGAG GGACCCATG AAGATTCAAC AAGTAGTTGT GAGAATTCCC TGTAATGGA
 16681 TGCTACCAAT TTGACATTG TAGACCTGCT ATTGTGTGCT TCTTATTGG GCTCTCCCAT
 16741 CTCCTTCACTT TCCAACCCAT ATTCCACATT AATCCCTTCC ACCACCATGC AACACTAGGT
 16801 AGGAGAGAAAG GAAGGTTAGA AGAGAAAAGTG GGTATAGATC TATTTAGACT ACTTCCTGCT
 16861 GATTAGGGGC AAGTCCAATC GTCAATTGTCA GGATACCTCC AACCAAGCAAC CAGCAAACCA
 16921 GCAAATCAGA AACAGCAAA GCAGCCAACA AGGCAGCACT AACCAAGCAGG ATTGGGGTCG
 16981 GTAGCGTGGG AGCAGTCACT ACTGGTCTTC TCATGGCTTT GGCATTAATA CTCTCTCAAG
 17041 AAATTCCGTA ATTTTTTCCC CACCCACTGA AATTCCGTAAT TTTAAATGC AACTATCTA
 17101 CAGCTGGCAA AAATCACATC TCTCCTAGAG CACAAGACAA ATCATAGTTA CTGGCTATTT
 17161 GCAATCTGAA GCATCTCAAT ATCCACACC TGGGATTAAA ACAAAACAT ATTCAACATCA
 17221 CATAACTGTT TTTTTTTCC AATTTTTTAT TAGGTATTTT CTTTATTTCAC ATTTCAAATG
 17281 CTATCCGAA AGTCCCCTAT ACCCTCCCCAC CTCCCTGCTC CCCTACACAC CCACTCCAC
 17341 TTTTGACCC TGGAGTTCCC CGGTACTGGG GCATATAAG TTTGCAAGAC CAAGGGGCCT
 17401 CTCTTCCCAG TGATGGCCGA CTAAGCCATC TTCTGCTACA TATGCAGATA GAGACACGAG
 17461 CTCTGGGGGT ACTAGTTAGT TCATATTGTT GTTCCACCTA TAGGGTCGCA GACCCCTTCA
 17521 GCTCCTGGG TACTTTGTCT AGCTCCTCCA CTGGGGGCTC TGTGTTTTAT CTAATAGATG
 17581 ACTGTGAGCA TCCACTCTG TATTGACAG GCACTGGCCT AGCGTCACAT GAGCCAGCTA
 17641 TATCAGGGTC CTTTCAGCAA AACCTTGCTG GCATGTGCAA TAGTGTCTGC GTTGGTGGT
 17701 TGATTATGGG ATGGATCCAC TAGTTCTAGA GC

//